TESTING
HUMAN LIMITS

AAAS

Science

SPECIAL ISSUE

TESTING HUMAN LIMITS

Next month, athletes from around the world will meet in Athens to push the limits of human performance in a variety of contests. This special section, with associated online material, explores the physiology, technology, and in some cases pharmacology that lie behind athletic feats. [Image: The Image Bank/Getty]

SPECIAL ONLINE CONTENT www.sciencemag.org/sciext/sports

science's next wave www.nextwave.org
CAREER RESOURCES FOR YOUNG SCIENTISTS

GLOBAL: Careers in Sports Science—From Treating the Masses to Supporting Athletes *E. Pain*

Next Wave's feature for August 2004 looks at career opportunities in sports science.

GLOBAL/UK: Finding the Right Track After Your Sports Science Degree *T. Kumana*

Many sports science graduates struggle to find challenging and stimulating work.

GLOBAL/US: Diary of a Sports Medicine Intern *R. Nicholls*

A former student of the American Sports Medicine Institute came all the way from Australia to learn about the biomechanics of baseball.

GLOBAL/US: My Journey Out From Under the Stairs *S. Seiler*

An American living in Norway allied sports and science with a Ph.D. in exercise physiology.

GLOBAL/CANADA: Talent Development in Science and Sports *N. Holt*

Research in talent development has some parallels to science careers.

GLOBAL/EUROPE: Keeping Your Finger on the Pulse—Sports Science in Europe *E. Pain*

Training and job opportunities in Europe, based on a report from the International Council of Sports Science and Physical Education in Berlin.

science's sage ke www.sageke.org
SCIENCE OF AGING KNOWLEDGE ENVIRONMENT

NEWS SYNTHESIS: Growing Pains *R. J. Davenport*

As researchers sort through the hype surrounding growth hormone, they hope to uncover the molecule's actual benefits.
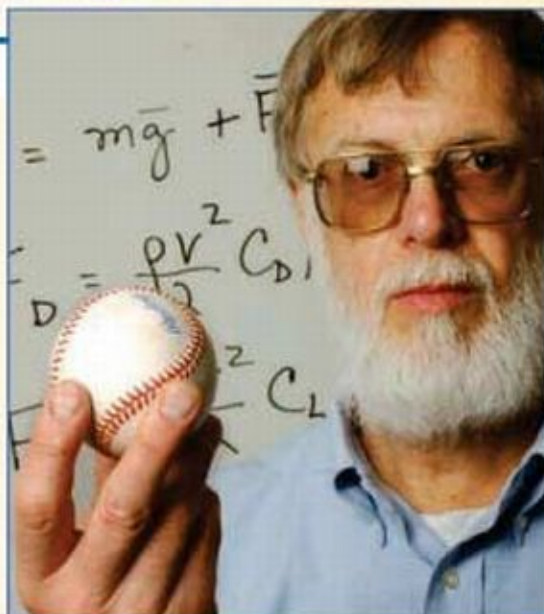
science's stke www.stke.org
SIGNAL TRANSDUCTION KNOWLEDGE ENVIRONMENT

PERSPECTIVE: Designer Androgens in Sport—When Too Much Is Never Enough *D. J. Handelsman*

Does anyone know the signaling properties of "designer" steroids?

REVIEW: Molecular and Cellular Determinants of Skeletal Muscle Atrophy and Hypertrophy *V. Sartorelli and M. Fulco*

IGF-1 is a crucial regulator of muscle mass.

Separate individual or institutional subscriptions to these products may be required for full-text access.

Science

586

613

618 & 665

Science

623 & 654

622 & 657

REPORTS CONTINUED

589
& 673

www.scienceonline.org

sciencenow www.sciencenow.org DAILY NEWS COVERAGE

More Muscle for Dystrophic Mice
Researchers make blood vessels leaky to deliver therapeutic genes.

Monster Waves Rule the Seas
Satellite survey reveals abundance of 25-meter-high waves.

Birds in the Fast Lane
Homing pigeons use highways to guide themselves home.

Pigeons follow roads.

science's sage ke www.sageke.org SCIENCE OF AGING KNOWLEDGE ENVIRONMENT

NEURODEGENERATIVE DISEASE CASE STUDY: Progressive Supranuclear Palsy *M. Pourfar and J.-P. Vonsattel*
An uncommon disorder has many features similar to those of Parkinson's disease.

NEWS FOCUS: High Anxiety *M. Leslie*
Youthful stress impairs adult brain's ability to mint new cells.

NEWS FOCUS: (Data)base Desires *R. J. Davenport*
New resource compiles genes that potentially influence human aging.

Strained and Stretched Nanoparticles

The electronic and optical properties of a material can change on going from bulk materials to the nanoscale. **Gilbert *et al.*** (p. 651, published online 1 July 2004) show how confinement effects can affect the bonding and packing of atoms. They use a number of techniques to measure the lattice structure and internal strains in 3-nanometer particles of zinc sulfide. A complex pattern of internal strains results as the particles attempt to lower the surface energy. These strains cause a reduction in the overall ordering and a stiffening of the lattice. When metals are deformed, the crystalline grains can rotate and realign, much in the way that painted shapes will stretch and warp when a canvas is pulled in a specific direction. **Shan *et al.*** (p. 654; see the Perspective by **Ma**) find that in nanocrystalline nickel, this type of deformation dominates, unlike the situation with coarser grained metals, where the production of grain boundary defects and dislocations accommodates most of the deformation energy. These results confirm many observations obtained from computer simulations and should help guide the design of optimum metals and alloys.

Lunar Meteorite Phones Home

A lunar meteorite found in the Sultanate of Oman (Sayh al Uhaymir 169) consists of four different impact breccias and is enriched in potassium, rare earth elements, and phosphorus. **Gnos *et al.*** (p. 657; see the Perspective by **Korotev**) used isotopic systematics to date the four impact events that occurred while the rock was at or near the surface of the Moon. The impact event dates of 3900 million years ago (Ma), 2800 Ma, 200 Ma, and <0.34 Ma, along with the chemical enrichments, help to pinpoint the source of the meteorite in the Lalande impact crater on the Moon. The dated impact events will allow lunar geologists to refine the ages of the different stratigraphic units associated with this meteorite into a more global model of the evolution of the Moon.

Removing Plant Defenses

In order to resist herbivore attack, plants use direct defenses, such as toxins and digestibility reducers, as well as indirect defenses that affect components of the plants' community (such as natural enemies and diseases). Plant defenses can be expressed constitutively or produced in response to an attacking pathogen or herbivore. **Kessler *et al.*** (p. 665, published online 1 July 2004; see the Perspective by **Dicke *et al.***) transformed the wild tobacco species *Nicotiana attenuata*, to silence three genes coding for enzymes in the jasmonate signaling pathway, which is known to be involved in induced plant defense. When planted into native habitats, the transformed plants were more vulnerable not only to their specialist herbivores but also to other herbivore species.

Pop Goes the Mitochondrion

In cells undergoing apoptosis or cell death, mitochondria, the powerhouses of the cell, often have a key role. Not only is cellular metabolism shut down, but mitochondria release molecules into the cytoplasm that further promote cell death. Substantial controversy has surrounded the mechanisms by which these processes occur. **Green and Kroemer** (p. 626) review the role of mitochondria in cell death. Permeabilization of the mitochondria can be the point-of-no-return that seals the fate of a cell, and numerous strategies are envisioned to alter these processes therapeutically to benefit patients suffering from a range of illnesses from cancer and heart failure to neurodegeneration.

Herbivores Drive Diversity

Habitat specialization and beta-diversity—the change in species composition between sites—may explain a large part of the overall diversity within tropical forests. However, why beta-diversity should be higher in the tropics remains unclear. To test the hypothesis that herbivores promote habitat specialization, **Fine *et al.*** (p. 663; see the Perspective by **Marquis**) performed reciprocal transplant experiments of specialist tree seedlings between soil types in the Peruvian Amazon, and also manipulated their herbivores. Habitat specialization of plants resulted from an interaction of herbivore pressure with soil type, which suggests that herbivores drive beta-diversity patterns by maintaining habitat specialization.

Controlling Interface Spin

The magneto-electronic properties of heterojunction structures formed between magnetic thin films and insulating layers are attractive for potential device applications. However, the behavior of such structures has been unpredictable, and techniques are needed that can investigate the influence of interface region. Using magnetization-induced second-harmonic generation, **Yamada *et al.*** (p. 646) show that they can probe and characterize the magnetization of buried interfaces formed between manganite and insulating thin films. Moreover, by grading the doping level of the interface region, they show how the properties can be altered in a controlled manner.
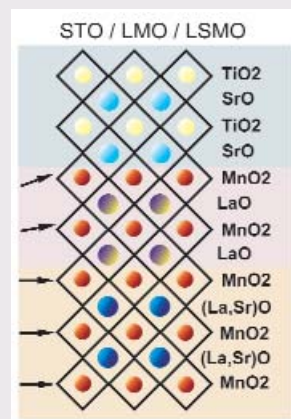
STO / LMO / LSMO

TiO2
SrO
TiO2
SrO
MnO2
LaO
MnO2
LaO
MnO2
(La,Sr)O
MnO2
(La,Sr)O
MnO2

Diffusion Goes Electronic

The atomic-scale resolution of the scanning tunneling microscope has been paired with the temporal resolution afforded by femtosecond laser pulses to differentiate electronically excited molecular diffusion from thermally induced diffusion. **Bartels *et al.*** (p. 648, published online 24 June 2004) excited CO molecules on the anisotropic Cu(110)

CREDITS: (TOP TO BOTTOM) YAMADA *ET AL.*; BARTELS *ET AL.*

surface with 200-femtosecond laser pulses at a wavelength of 405 nanometers. Unlike diffusion at thermal or equilibrium conditions, which occurs along the rows of atoms formed by Cu atoms on this surface, the nonequilibrated electronically excited CO molecules diffused over the rows as well. A phenomenological model can account for these results in terms of electronic excitation of the CO-substrate vibrations.

Worming into Whale Bones

A new genus of annelid worm that is related to hydrothermal vent worms has been discovered on the corpse of a gray whale found several thousands meters deep off the coast of California. **Rouse et al.** (p. 668) have named the genus *Osedax*. The female worms possess tubes from which red plumes emerge and which harbor numerous, non-feeding, dwarf male worms. Like vent worms, *Osedax* worms are gutless and contain bacterial symbionts. The worms burrow into the whale bones and form rootlike structures which contain the symbiotic organotrophic bacteria that mobilize nutrients from the whale skeleton.

The Genomics Underlying Acne

Propionibacterium acnes is a ubiquitous, human skin–dwelling organism involved in the etiology of acne. **Brüggemann et al.** (p. 671) have sequenced and analyzed the complete genome of *P. acnes*. The genome data offer information on the bacterial antigens and tissue-damaging enzymes that may cause the inflammatory reactions underlying the disease process.

Freeze-Frames of Motor Movement
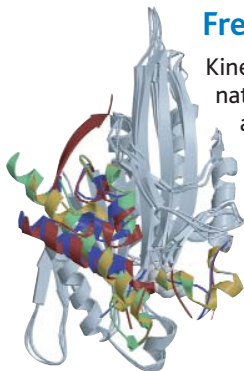
Kinesin motor proteins move along microtubules by rapidly alternating between tightly bound and detached states. Movement is adenosine triphosphate (ATP)–dependent and changes in binding affinity are associated with the ATPase cycle. **Nitta et al.** (p. 678) report crystal structures of the monomeric kinesin KIF1A with three transition state analogs. Kinesin alternately uses two loops to bind microtubules with an intermediate state in which neither loop binds. When KIF1A is working, it likely alternates between a tight-binding state with the affinity biased toward the forward tubulin subunit, and a weak-binding state that allows one-dimensional diffusion.

Reconstituting Prion Disease in Mice

The prion hypothesis postulates the existence of infectious proteins capable of propagating disease. **Legname et al.** (p. 673; see the news story by **Couzin**) now present evidence that a novel strain of prion disease can be induced in mice injected with recombinant prion proteins. Brain extracts from these mice could then be used to infect other mice to cause a neuropathological disorder distinct from other known strains of prion disease.

Host-Parasite Gene Transfer in Plants

The parasitic plant family Rafflesiaceae resisted definitive taxonomic placement since its initial description nearly two centuries ago. Recently, a study placed it firmly in the order Malpighiales, based on the mitochondrial gene *matR*. **Davis and Wurdack** (p. 676, published online 15 July 2004) have reexamined this question by adding all family representatives of Malpighiales across four genetic loci spanning the nuclear and mitochondrial genomes. The nuclear DNA, and one mitochondrial locus, confirmed the position of Rafflesiaceae within Malpighiales. However, the other mitochondrial locus, *nad1B-C*, places Rafflesiaceae in the family Vitaceae, which is in a different order. These incompatible phylogenetic results appear to provide a new example of horizontal gene transfer between species—horizontal gene transfer mediated by a plant host-parasite system.

CREDIT: NITTA *ET AL.*

EDITORIAL

Here Come the Olympics

The Olympics are upon us, and we sports fans can't wait for the nonstop television diet of sports molded from the classic Greek tradition, like wrestling and track, and of others added to the occasion, like rhythmic gymnastics, synchronized swimming, and even baseball. Political pressure in support of a primarily national sport can put it on the list, just as can the appeal of the sport itself. And the Olympics have shown little capacity to resist either the proliferation of new sports or the professionalization of old ones. How long, one wonders, will they be able to hold out against the "extreme sports" categories now represented in the X Games?

What drives this diversification is partly revenue, which of course means television. But there are other forces that have much to do with science. The Olympics selects athletes performing at the edge of their physical capacity, pushing competitors into training regimes unheard of decades ago. The same urge has moved the Olympic Committees toward accepting professional athletes as competitors, which surely has the sometime Olympic czar Avery Brundage revolving in his grave. This oddly sporadic surge toward professional acceptance yields a perplexing heterogeneity of treatment: "Dream teams" of National Basketball Association players are dispatched to represent the U.S. and set up in luxury hotels, while America's college baseball players bunk with the rest of the plebeians.

Here's another science-based change in the Games: it's how materials science has transformed some of the traditional sports. I actually can remember the first 15-foot pole vault, but after the properties of fiberglass converted the vaulter's instrument from a pole to a catapult, we entered a new record-setting domain. In cycling and yachting, technology probably accounts for more of the variance in outcome than in other Olympic sports. (A friend of mine resents this, refusing to take seriously any sport that depends on the device as well as the athlete.)

The big science problem, though, is that in the sports that most directly measure individual athletic ability, there is no guarantee that the playing field is level. Drug violations are not new to the Games; some winning distance runners were charged with blood doping decades ago, and more recently the Canadian sprinter Ben Johnson was stripped of his medal because of steroid abuse. Now several U.S. track and field athletes, including a few prospective Olympic competitors, are under suspicion, and others will remain radioactive until the testing regime improves enough to earn public trust. What we now have is a pharmacological arms race between the detection technology of the anti-dopers and the inventiveness of the designer-steroid mavens. It is a close contest, and if past is prologue, we cannot know who is ahead at any given moment. I liked track and field a lot more before they took it into the lab.

Fortunately, the really significant performance gains have come not from drugs but from better understanding of the body's limits and the role of training in overcoming them. Dr. Roger Bannister used elegant experiments on his own respiratory physiology to help shatter a record once thought unbreakable. Now dozens of runners from around the world can beat his time by 15 seconds or so. And there has been a remarkable change in our ideas about what women can do in events previously dominated by men, exemplified by the rapid convergence of the women's times in distance races toward the best men's times.

Science surely has had a mixed impact on the Games: It has been used to enhance human capacity through improved training and better technology, but it has also brought us clever ways to cheat. As for me, even though I know that everything may not be on the level, I really am looking forward to the Olympics. So let the Games begin! I plan to adopt the English poet Samuel Taylor Coleridge's advice and follow the events having willingly suspended disbelief, confident that the playing field is level, that no one is on drugs, and that no athlete has a concealed bionic assist. Don't laugh; it works for me.
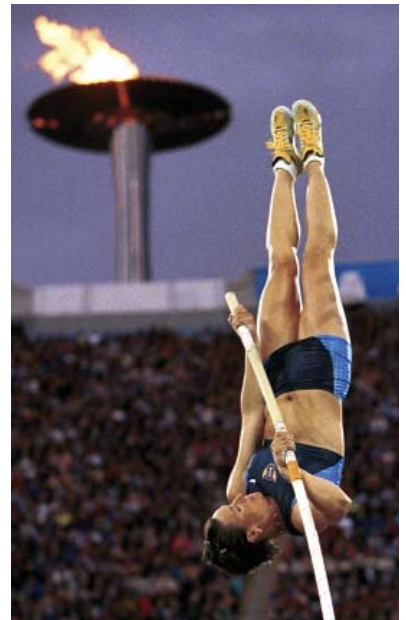
Donald Kennedy
Editor-in Chief

NEWS

This Week
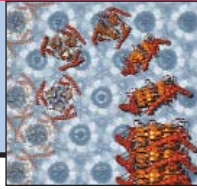
PAGE 589
Prion proof?
Perhaps

591
Contention
continues over
Kennewick
bones

THEORETICAL PHYSICS

Hawking Slays His Own Paradox, But Colleagues Are Wary

DUBLIN, IRELAND—In a public appearance that drew worldwide media coverage, Stephen Hawking claimed last week that he had solved one of the most important problems in physics: whether black holes destroy the information they swallow. Speaking at a conference here[*] in a lecture hall packed with physicists and reporters, the University of Cambridge professor reversed his long-standing position and argued that information survives. As a result, Hawking conceded the most famous wager in physics and handed over an encyclopedia to the winner of the bet.

"It is great to solve a problem that has been troubling me for nearly 30 years," Hawking said during his presentation. Other physicists, however, doubt that Hawking has solved the long-lived puzzle. "It doesn't seem to me to be convincing," says John Friedman, a physicist at the University of Wisconsin, Milwaukee.

The question of what happens to information when it falls into a black hole goes to the heart of a central idea in modern physics. Just as scientists in the 19th century figured out that energy can be neither created nor destroyed, many 20th century physicists concluded that information is also conserved. If true, information conservation would be one of the most important principles in science—perhaps more profound even than conservation of mass and energy. Unfortunately, there was a big obstacle: black holes.

When an object falls into a black hole, its mass and energy leave an observable imprint by making the black hole more massive. According to general relativity, however, any information the object carries is irretrievably lost: An outside observer couldn't tell whether the black hole had swallowed a ton of lead, a ton of feathers, or a ton of Ford Pintos. If black holes can destroy information in this way, information conservation cannot be a universal law.

The debate raged for decades whether black holes were an incurable exception to

the permanence of information. In the 1970s, Hawking and some of his colleagues, including Kip Thorne of the California Institute of Technology (Caltech) in Pasadena, argued that black holes trump information. Others, such as Caltech's John Preskill, argued that some undiscovered loophole would keep information safe until the black hole somehow disgorged it. In 1997, Hawking and Thorne made a wager with Preskill;

Not proven? Stephen Hawking's new view of black holes rests on unusual math.

the winner was to receive an encyclopedia of his choice, from which information can always be retrieved.

At the Dublin conference, Hawking conceded the bet. Using a mathematical technique known as the Euclidean path integral method, Hawking proved to his own satisfaction that information is not, in fact, destroyed when it falls into a black hole. "If you jump into a black hole, your mass-

energy will be returned to our universe … in a mangled form which contains the information about what you were like, but in a state where it cannot be easily recognized," said Hawking. That implies that black holes are not portals to other universes, a possibility Hawking himself had suggested. "I'm sorry to disappoint science-fiction fans," he said.

In conceding the bet, Hawking presented Preskill with *Total Baseball: The Ultimate Baseball Encyclopedia*. Thorne, however, refused to admit defeat. "I have chosen not to concede because I want to see more detail," he said, but added, "I think that Stephen is very likely right."

Others are less certain. Friedman, for one, has doubts about Hawking's mathematical method. Quantum field theorists are happy to use the Euclidean path integral technique for problems involving particles and fields, but most gravitational theorists avoid it because it produces equations riddled with hard-to-reconcile infinities. They prefer a more straightforward "Lorentzian" approach to gravity. Nobody has proven that the two methods always give the same results. "I'm skeptical whether the Euclidean path integral method generally represents the evolution of spacetime that is really Lorentzian," says Friedman. If not, then Hawking's conclusion may be an artifact of the mathematical method rather than a general result. Another reason for skepticism, Friedman says, is that Hawking's calculation takes a sum over all possible idealized black hole locations and all observers in the universe, but the results don't seem to apply to a specific black hole and a specific observer.

In part because of the Euclidean method, Hawking's work doesn't seem to yield any insight into how black holes preserve or release information—whether all the pent-up information bursts forth at once, or whether it trickles out as subtle correlations in radiation coming from the black hole. Even Preskill says he wishes that Hawking's argument made more physical sense and could be expressed in more conventional mathematical terms. "If one could extract from the calculation an understanding that could be reproduced in a purely Lorentzian calculation, that would help a lot," he says.

Despite his doubts, Preskill has no qualms about accepting *Total Baseball*. "The terms were that the winner would receive the encyclopedia when the other party concedes," he says. "I don't have to agree."

–CHARLES SEIFE

Focus

594
The ultimate flu experiment
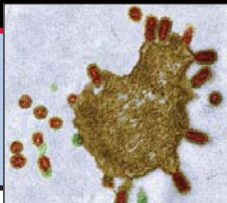
596
Neuroscience and the death penalty
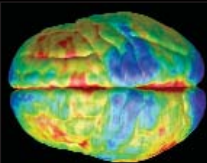
600
Insights from an outburst
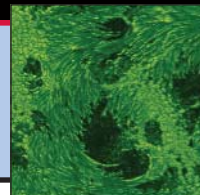
U.S. SCIENCE BUDGET

Caught in a Squeeze Between Tax Cuts and Military Spending

Banking on the benevolence of a lame-duck Congress is risky business. But for U.S. scientists, a possible postelection session may be the best bet to salvage research programs that are facing budget cuts.

Last week, Congress began a 6-week break having finished work on only one of the 13 spending bills for the 2005 fiscal year that begins on 1 October. Although the lone completed bill provides modest increases for defense research, the House has taken some initial steps on a dozen other agencies that suggest most science programs are in for a very hard time this year. The National Science Foundation (NSF) and NASA are facing real cuts, and the National Institutes of Health may have to settle for a small increase that may not keep pace with inflation (*Science*, 16 July, p. 321). The dark budget may hold silver linings for the Department of Energy's (DOE's) science programs and the National Institute of Standards and Technology (NIST). But even those gains could be at risk once Congress returns in September for a last-gasp attempt to finish its fiscal business before the November elections.

The squeeze is a result of Republican-led efforts to reduce taxes and hold down domestic spending while fighting wars in Iraq and Afghanistan and defending against terrorism at home. That has left the 13 spending panels that divvy up the government's $2 trillion budget with less money than agencies requested. The latest bad news came on 22 July when a House panel voted to shrink the budgets of NSF and NASA by 2.1% and 1.5%, respectively, below this year's levels.

The decline for NSF, which would be the first in nearly 2 decades, contrasts with a 3% increase requested by President George W. Bush. It also makes a mockery of a 15% annual rise called for by a 2001 law that, unfortunately for scientists, appropriators don't have to follow. "We're still hopeful that the numbers will improve after the Senate has acted," says acting NSF Director Arden Bement. "We are dealing with some frustrated appropriators."

The frustration stems from the fact that NSF and NASA are part of a larger spending bill that also funds the Veterans Administration. Historically, veterans' needs take precedence, especially in an election year. This year, the House panel approved an extra $2.5 billion for veterans' health care, leaving little new money for other agencies. "We can't compete with the veterans," says Sam Rankin, chair of the Coalition for National Science Funding and a lobbyist for the American Mathematical Society.

The House bill would trim NSF's bread-and-butter research programs by $109 million, to $4.1 billion. It would cut the $935 million education directorate by 10%, in-

2005 Science Budgets So Far

Bleak house. Congress has approved only the DOD budget; funding bills for other agencies are awaiting full House and Senate action.

cluding no new funding for a program that links universities with local schools. It would also delay the start of the National Ecological Observatory Network while allowing design work on two prototype sites.

For NASA, the $228 million cut reverses a Bush Administration proposal for a $1.1 billion increase for moon and Mars exploration. Several new space science missions took it on the chin while the committee piled on millions of dollars in earmarks. The panel rejected the entire $70 million requested to begin a robotic lunar exploration effort and nixed $12.4 million to start the scientific work on a Jupiter Icy Moons Orbiter and nearly all of the $17.6 million proposed for an Orbiting Carbon Observatory. At the same time, the panel added goodies such as $150,000 for the Coca Cola Space Science Center in Columbus, Georgia, and $3 million for the National Center of Excellence in Bioinformatics in Buffalo, New York.

The committee also refused to fund a new Crew Exploration Vehicle that ultimately could send humans to the moon and Mars. But it fully funded the $4.3 billion request for the space shuttle. The panel said it backed the idea of Bush's exploration vision but noted that the committee "does not have sufficient resources." The White House says it may veto the bill if the NASA numbers don't improve.

One agency that took a big hit last year may get a chance to climb partway out of its budget hole. On 8 July, the House approved an 11% increase for the intramural programs at NIST and told the agency to spend whatever it takes from its research account to outfit its new Advanced Materials Laboratory. Eighty-two NIST employees have accepted buyouts, and a better 2005 budget, says Bement, who also heads NIST, means that "we won't have to lay off any scientists."

At DOE, science advocates are praising a 25 June House vote giving the agency's science office a 3% boost to $3.6 billion, rejecting a White House call for a modest cut. Supercomputing research was a big winner, getting a 16% jump to $234 million. DOE's heavily earmarked biological research account, however, would slump 11% to $572 million. Among the victims is a new $5 million molecular tag production facility. Lawmakers said they didn't like the department's plan to allow only DOE labs to bid for the project.

Defense researchers are pleased with an 8% increase, to $1.5 billion, for basic research included in a Department of Defense (DOD) spending bill to be signed soon by Bush. That reverses a proposed 5% cut. Applied research would jump 12% to nearly $5 billion.

Now, science advocates are waiting to see how the Senate deals with other science budgets. But many observers predict that the final numbers won't be settled until late this year, in a lame-duck session after the elections.   **–JEFFREY MERVIS**

With reporting by David Malakoff and Andrew Lawler.

BIOMEDICINE

An End to the Prion Debate? Don't Count on It

A bold set of prion experiments in mice may have proven that the misshapen proteins are, by themselves, infectious. If the work holds up, it will be a watershed in prion biology, validating the belief that these proteins alone are the culprits in "mad cow disease" and similar illnesses.

As is typical for the controversy-laden field, however, many scientists express reservations about the study on page 673. It was led by Stanley Prusiner of the University of California, San Francisco, who won the Nobel Prize in 1997 for discovering prions.

"It's really a striking result that seems to
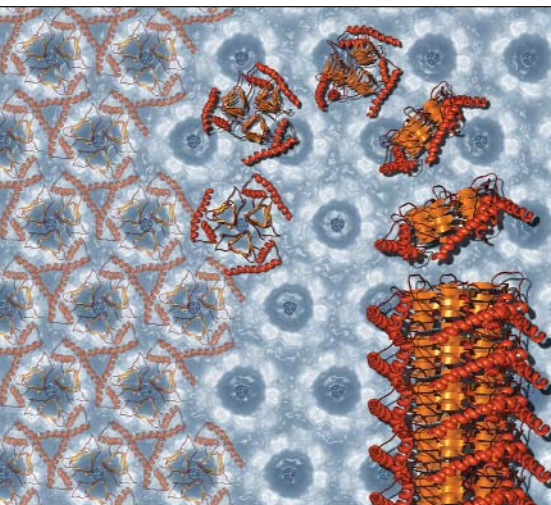
Building a prion? In a model of prion formation, misshapen PrP proteins (red) stack up into amyloid fibrils.

fill in one more piece of the infectivity puzzle," says Byron Caughey, a biochemist at the National Institutes of Health's Rocky Mountain Laboratories in Hamilton, Montana. "But," he adds, "it's worth pointing out some significant caveats."

For years, biologists have tried to prove that a protein called PrP can misfold and become an infectious prion by purifying protein clumps from diseased brains and injecting them into healthy animals. But it hasn't been clear that PrP alone was what was being injected; using synthetic misfolded PrP, meanwhile, hasn't reliably triggered disease.

In their tests, Prusiner and colleagues used transgenic mice making 16 times the normal amount of PrP. These mice express a truncated PrP that may more readily make up prion clumps. This, the group reasoned, might sensitize the animals to introduced PrP.

To obtain PrP free of brain tissue, Prusin-

er's team genetically altered *Escherichia coli* bacteria into producing PrP fragments that they misfolded to form amyloid fibrils, which have been implicated in various brain diseases. Prusiner's team injected those prion fibrils into the brains of the mice.

Almost a year later, with no animals sick, the researchers were ready to declare the study a failure. But then, 380 days after being inoculated, one of the mice showed symptoms of a prionlike disease. Eventually, all seven inoculated mice showed neurological disease, the last one 660 days after injection.

Prusiner's team also inoculated a batch of normal animals with brain tissue from one of the sick ones. These rodents took about 150 days to sicken.

"It is a spectacular breakthrough," says Neil Cashman, a neuroscientist at the University of Toronto. "This is the beginning of the end of all the objections about the prion hypothesis."

Not so fast, say some experts. Do Prusiner's mice with excess PrP get sick normally? wonders John Collinge, director of the Medical Research Council Prion Unit at University College London. His team had relied on rodents with 10 times the normal level of PrP but abandoned them after finding prion disease–like pathology in animals that hadn't been inoculated with anything.

Prusiner's mice, says Collinge, may be "poised" to become infectious even without the inoculation; giving them a shot of synthetic, misfolded PrP may push them over the edge, but so might other stresses.

The long latency time between inoculation and disease also worries prion experts. Some wonder if the experiments were contaminated by other prion strains in the lab. Yale University neuropathologist Laura Manuelidis, who has long criticized the prion hypothesis, says the brain samples from some of Prusiner's mice resemble RML scrapie, a common strain. Prusiner counters that with contamination, the control animals inoculated with saline should have gotten sick as well.

Another explanation for long latency is that infecting animals with synthetic PrP is inefficient. The first inoculations may have contained few prions, says Prusiner. This might also explain why no one has yet accomplished the gold-standard experiment: infecting normal mice, not transgenic ones, with pure prion proteins. Until then, one of biomedicine's longest-running controversies is likely to continue. **–JENNIFER COUZIN**

CREDIT: G. LEGNAME *ET AL.*

ScienceScope

GAO Faults Science Agencies on Title IX Compliance

Three top U.S. science agencies have failed to enforce a federal law aimed at increasing female participation in educational programs, according to a report unveiled last week by the Government Accountability Office (GAO), Congress's investigative arm. The report, which was requested by Senators Ron Wyden (D–OR) and Barbara Boxer (D–CA) after a 2002 hearing (*Science*, 11 October 2002, p. 356), says the Department of Energy, NASA, and the National Science Foundation have not been monitoring grantee institutions to check if they are complying with Title IX. The 32-year-old legislation, which allows the government to withhold funds from institutions that practice gender discrimination, applies to all fields of education, but its impact has mostly been limited to athletics.

The GAO report confirms that the federal government needs to enforce Title IX "not just on the playing field but also in the classroom," says Wyden. He believes compliance reviews by granting agencies are essential to close the gender gap in the sciences and engineering. Massachusetts Institute of Technology biologist Nancy Hopkins, who chaired a study on the status of women faculty members at MIT's School of Science, predicts that "without government oversight and support, the full participation of women and minorities in science and engineering will not occur in our lifetime—or in the lifetime of our children."

–YUDHIJIT BHATTACHARJEE

Mexico Approves Genomic Medicine Institute

After 5 years of discussion, Mexico is getting a new institute for genomic medicine. President Vicente Fox last week approved construction of the $200 million INMEGEN center in Mexico City, which is expected to employ 120 researchers and open it first units next year.

The institute, which will focus in part on disease susceptibilities among Mexico's dozens of indigenous groups, will be led by biomedical researcher Gerardo Jiménez-Sánchez of Johns Hopkins University in Baltimore, Maryland (*Science*, 11 April 2003, p. 295).

"We cannot afford the luxury of not joining this knowledge revolution," said Fox. Jiménez-Sánchez says INMEGEN researchers will not work with human embryos. Mexican law allows both embryo research and therapeutic cloning.

–XAVIER BOSCH

ARCHAEOLOGY

Wisconsin Dig Seeks to Confirm Pre-Clovis Americans

MILWAUKEE, WISCONSIN—This week, archaeologists will begin to dig 48 kilometers south of here, at a site that even skeptics say may be the most convincing yet in demonstrating the early presence of humans in the Americas. Scientists will search a mucky lakeside just west of the city of Kenosha for additional remains of a woolly mammoth. Bones found previously bear marks from human butchering and have been dated to 13,500 radiocarbon years before present—a full 2000 years before big-game hunters known as the Clovis people were thought to have arrived on the continent.

Sites near Kenosha "may be the best pre-Clovis sites in North America," says team leader Michael Waters of Texas A&M University in College Station. Even pre-Clovis skeptic Stuart Fiedel, an archaeologist with the Louis Berger Group in Washington, D.C., agrees that "the Kenosha sites are high up on my radar screen. On the face of it, they seem to be one of the best cases [of pre-Clovis evidence]."

Archaeologists long thought that America was first settled by the Clovis hunters, who crossed the Bering Strait and moved south through an ice-free corridor around 11,500 radiocarbon years ago. Then in recent years dozens of sites in both North and South America pointed to an even older human occupation. But each pre-Clovis site

Mammoth meal? Bones from Kenosha, dated to 12,500 radiocarbon years ago, show signs of butchery by early Americans.

has been bitterly contested (*Science*, 2 March 2001, p. 1730), and a handful of influential archaeologists believes that definitive pre-Clovis evidence is lacking. "One of my problems with the [pre-Clovis] position is that the sites that it is founded on are still

dubious," says Fiedel.

Hence the excitement over the sites near Kenosha. In 1990, an amateur archaeologist found butcher marks on mammoth bones stored at a local historical museum; archaeologists later excavated at two sites, those of the Schaefer and Hebior mammoths. These mammoth bones are so well preserved that collagen could be extracted from inside the bone for radiocarbon dating, yielding dates of about 12,500 radiocarbon years ago, 1000 years before the Clovis people. And a handful of crude stone tools—unlike the elegant spear points of the Clovis people—were recovered under the bone piles. All in all, the sites are unique, with "unequivocal stone tools [and] excellent dates," says Waters.

Now his team is in pursuit of an even older Kenosha mammoth at Mud Lake, where a few bones with cut marks were unearthed during a ditch-digging project in the 1930s and later dated. Waters believes that the rest of the mammoth is there and plans to try to relocate it this summer while scouting for new sites for future excavations. The preliminary dig starts this week, but because heavy rains have slowed the work, full-scale excavation of Mud Lake isn't expected until next year.

Given the potential of the Kenosha sites, they have attracted little attention. "I really don't understand why there has not been more investigation devoted to [them] to date," says Fiedel. Starting this summer, Waters's crew hopes to change that.

–**TERRENCE FALK**

Terrence Falk writes on science, education, and public policy from Milwaukee.

ENVIRONMENT

States Sue Over Global Warming

In a legal gambit aimed against global warming, the attorneys general of eight states last week sued the five largest emitters of carbon dioxide in the United States for creating a public nuisance. The states are asking that the electric utility companies cut emissions by 3% each year for a decade. Legal experts predict the states' case will be an uphill battle.

Carbon dioxide litigation is heating up. In 2002, environmental groups sued the Overseas Private Investment Corp. and the Export-Import Bank of the United States for not conducting environmental reviews on the power plants they financed. And last year, Maine, Massachusetts, and Connecticut sued the Environmental Protection Agency for not regulating CO_2 as a pollutant under the Clean Air Act. Now, the states have taken the first legal action directly against CO_2 emitters.

The plaintiffs—California, Connecticut, Iowa, New Jersey, New York, Rhode Island, Vermont, and Wisconsin, along with the

City of New York—claim that the CO_2 that utility companies release contributes to global warming, which will harm state residents. The alleged ills include increased numbers of deaths from heat waves, more asthma from smog, beach erosion, contamination of groundwater from rising sea level, and more droughts and floods. "The harm to our states is increasing daily," Eliot Spitzer, the attorney general of New York state, said at a press conference.

The defendants together spew about 650 million tons of CO_2 a year. Their 174 fossil fuel–burning plants contribute roughly 10% of the anthropogenic CO_2 in the United States. The suit maintains that annual cuts of 3% are feasible through making plants more efficient, promoting conservation, and using wind and solar power—without substantially raising electric bills. "All that is now lacking is action," Spitzer said.

That claim irks American Electric Power

of Columbus, Ohio, a defendant. Spokesperson Melissa McHenry says that the company had already committed to reducing its emissions by 10% by 2006. "Filing lawsuits is not constructive," she says. "It's a global issue that can't be addressed by a small group of companies."

It will also be a tough suit to win, says Richard Brooks of Vermont Law School in South Royalton, who studies the legal issues of air pollution. The fact that global warming is a planetwide phenomenon will make it difficult to establish how much these companies are contributing to the claimed harm. And under public-nuisance law, the plaintiffs must show that their citizens are suffering significantly more than the nation as a whole. "I would be totally amazed if the court gave this a serious response," Brooks says. "This makes me imagine that this is more of a symbolic suit."

–**ERIK STOKSTAD**

Court Battle Ends, Bones Still Off-Limits

When Native American tribes decided last week not to fight an appeals court ruling, it looked as though the way was clear for scientists to study the 9300-year-old skeleton called Kennewick Man, which has been tied up in legal battles for the past 8 years. But scientists say that although the ruling sets a favorable precedent for studying other ancient skeletons, they are not optimistic about getting to study Kennewick Man himself anytime soon. The government continues to find fault with outside scientists' research plans and to deny access to the remains. Negotiations are in progress, but the lawyer for the eight scientist-plaintiffs in the suit, Alan Schneider of Portland, Oregon, says, "we are still far apart." Going back to court, he adds, "is definitely a possibility."

The Kennewick case finally appeared to have come to an end on 19 July when the defendants, four tribal groups, decided not to appeal to the U.S. Supreme Court a decision by the 9th U.S. Circuit Court of Appeals. That court ruled that because there is no evidence linking the Kennewick skeleton to any existing tribe, the Native American Graves Protection and Repatriation Act (NAGPRA) does not apply to it (*Science*, 13 February, p. 943). The court's interpretation of NAGPRA is a significant advance that will have "major implications" in other cases in which Native American groups are claiming remains, says Robson Bonnichsen of The Center for the Study of the First Americans at Texas A&M University in College Station. In a U.S. Army Corps of Engineers project in Texas, for example, he says, Native Americans at first claimed remains from a 4000-year-old burial ground, but a compromise has been reached so that scientists will have access to them.

Meanwhile, scientists are eager to study Kennewick Man, one of the oldest skeletons in North America. Schneider says that in 2002, the scientists submitted a 40-page study plan to the Department of the Interior and the Corps of Engineers, which has custody of the remains at the Burke Museum in Seattle. It is "a state-of-the-art proposal to do the most detailed look at a first American that has ever been put together," says Bonnichsen. "We wanted to do a class act."

But officials at Interior and the Corps of Engineers have responded with a throng of objections. According to Bonnichsen, the Corps of Engineers says the skeleton is "fragile," and officials seek to limit the number of scientists who have access to it. "The corps is concerned about the condition and wants to limit handling to what is needed to produce new knowledge," says Frank McManamon, chief archaeologist at the National Park Service. McManamon, who has been advising on the government response to the study plan, says the plan doesn't "build on the substantial amount of scientific investigation that has already been done" by government-appointed scientists. For example, he says that Bonnichsen and colleagues want to take bone samples for DNA testing, even though sampling has already been done and three separate labs couldn't extract any DNA.

Lawyer Schneider counters that the government-sponsored radiocarbon and DNA tests "used or damaged up to 60 grams of the skeleton," whereas the scientists have proposed "microsampling," which would destroy no more than 1.5 grams of bone. He adds that many other areas need study. For example, although government-appointed scientists did computed tomogra-
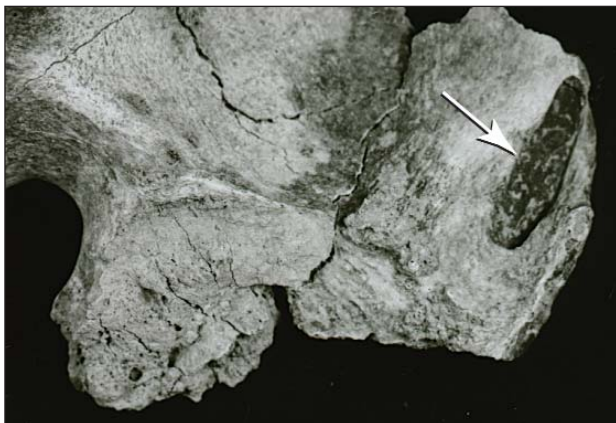
Still fighting. Scientists seek access to the bones of Kennewick Man, who died with a projectile point in his pelvis (arrow).

phy (CT) scans to examine the projectile point lodged in the skeleton's pelvis, Schneider says that "there is still a major controversy over which direction [it] entered," and that more sophisticated CT technology is now available to study it. "What Frank [McManamon] seems to be saying is 'We've looked at them, so you don't need to' "— hardly a scientific stance, says Schneider.

While the haggling continues, Native Americans have indicated that they will now embark on a nationwide campaign to pressure Congress to rewrite NAGPRA.

–**CONSTANCE HOLDEN**

ScienceScope

Panel Pans UC-Novartis Deal

The University of California (UC), Berkeley, should pass on any proposals similar to the agreement it once made with pharmaceutical giant Novartis, according to a new independent report commissioned by the school's academic senate. In 1998, Novartis pledged $25 million over 5 years to the plant and microbial biology department in exchange for significant access to the department's labs and scientific discoveries. The agreement was greeted with outrage by many researchers at the school and across the country (*Science*, 17 January 2003, p. 330).

The direct impacts of the pact on the university "have been minimal," concludes the report, authored by food and agricultural specialist Lawrence Busch of Michigan State University in East Lansing and colleagues. Although graduate students in the field enjoyed increased stipends, "few or no benefits" in terms of patent rights or income went to the university or to Novartis and its successor Syngenta, according to the report. Busch and his co-authors also concluded that the agreement did not damage the department's basic science efforts, as many opponents feared it would. But the report recommends that Berkeley avoid future industry agreements "that involve complete academic units or large groups of researchers" and urges "broad debate early in the process of developing new research agendas." The study will be submitted to the Berkeley Senate on 1 August for consideration.   –**ANDREW LAWLER**

House Cuts EPA R&D, Restores STAR Grants

Research budgets at the Environmental Protection Agency (EPA) would face a 4.3% cut, to $589 million, in a spending plan approved last week by the House appropriations committee. The cuts are part of an increasingly gloomy budget picture for science (see p. 587).

Environmental researchers did get some good news, however. The panel restored funding to EPA's extramural grants program, called Science to Achieve Results (STAR), bringing the program back to its fiscal year 2004 level of about $76.1 million, with an additional $9.5 million spent on graduate fellowships. In February, the Bush Administration proposed deep cuts to both STAR grants and the fellowships. The House support for STAR is encouraging, says Craig Schiffries, director of science policy at the National Council for Science and the Environment in Washington, D.C., but he's disappointed that the funding is still lower than the $100 million requested in recent years.   –**ERIK STOKSTAD**

Congressmen Clash on Politics and Scientific Advisory Committees

For the past several months, the Bush Administration has responded with strong denials to charges that it has chosen members of scientific advisory committees in part for their political views. The charges are either wrong or distorted or they reflect aberrations in the selection process, Administration officials have asserted (*Science*, 16 July, p. 323). But last week a prominent House member took a different tack: There's nothing wrong with mixing science and politics in determining the makeup of scientific advisory committees, says Representative Vern Ehlers (R–MI).

Ehlers, a former physics professor and staunch conservative, offered this view in an impromptu debate with Representative Henry Waxman (D–CA), a dyed-in-the-wool liberal, at a meeting of the National Academies' Committee on Science, Engineering, and Public Policy. The committee is taking its third stab at recommending how the government can improve its pool of scientific and technology talent. Its previous reports focused on ways to make full-time jobs in Washington, D.C., more welcoming to scientists, but this year's effort is also examining the hundreds of outside advisory committees that help federal agencies do their work. The panel, which includes veterans from previous administrations spanning both parties, hopes to deliver its report soon after the November election.

The problems flagged in earlier reports still exist: an intrusive and time-consuming vetting process, a likely cut in salary, and the uncertainty of winning Senate confirmation. Panel chair John Edward Porter, a former representative from Illinois and patron of the National Institutes of Health, says the issues remain "intractable." But Porter's first question to his former colleagues signaled that, this time around, the burning questions are more political than logistical. "Do you think that it's appropriate for the government to ask someone being considered for an advisory position, 'Who did you vote for?'" Porter wanted to know.

"I think that it's an appropriate question to ask," replied Ehlers, who also defended the practice of asking where potential advisers stand on various hot-button issues. "Abortion is not a scientific issue, and yet there are technical committees that give advice on issues relating to abortion, like the use of embryonic stem cells in research," he
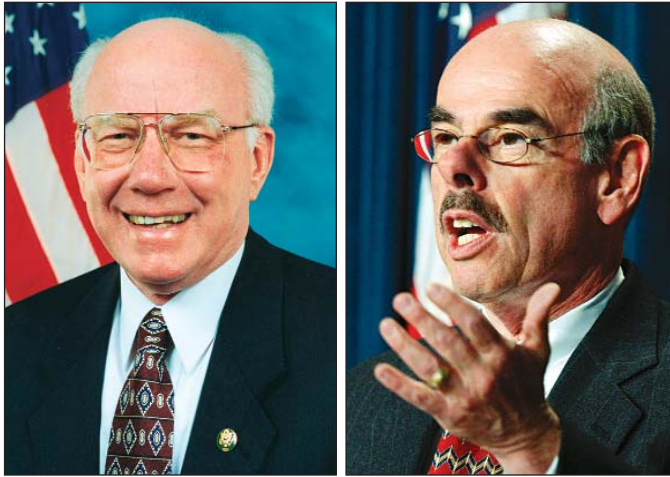
En garde. Vern Ehlers (*left*) says it is appropriate to ask potential panel members whom they voted for; Henry Waxman disagrees.

said. "The dividing line [between politics and science] is not clear. My first principle is to make sure that all views are represented at the table, to get the best people, and then let them shout at each other. That's the ideal scientific advisory committee."

Waxman rejected that argument. "There is a line you need to draw," he insisted. "For political appointees, the president should expect that his nominee supports his policies. But for advisory committees, they ought not to ask one's views on abortion, or how they voted [in the 2000 presidential election]." Waxman later insisted that the Bush Administration has imposed its own judgments on the advisory process, "settling on a policy first and then finding scientists to support that view."

Earlier in the day, presidential science adviser John Marburger told the panel that the candidate "pool is alarmingly small" when it comes to hiring good federal science managers. But he dodged a question from one of his forerunners, Frank Press, about interference from the White House in staffing his Office of Science and Technology Policy. Resisting such intrusions, Marburger said, "is easier than you might think."

–JEFFREY MERVIS

Los Alamos Suspends 19 Employees

The Department of Energy's (DOE's) security and safety problems continue to escalate. George "Pete" Nanos, head of Los Alamos National Laboratory in New Mexico, last week suspended 19 employees—including some senior scientists—pending an investigation of possible rules violations. He had already shut down virtually all work at the lab until the investigation is completed (*Science*, 23 July, p. 462). Then, starting this week, DOE Secretary Spencer Abraham suspended classified work involving portable computer disks at all DOE facilities, including Lawrence Livermore National Laboratory in California.

The massive "stand down" is needed, Abraham says, to make sure that security lapses at Los Alamos weren't repeated elsewhere and to "make certain that specific individuals can be held responsible and accountable for future problems."

Both moves are rooted in a 7 July inventory at Los Alamos that concluded that two classified disks were improperly removed from a safe. Then, on 14 July, an intern's eye was injured by a research laser that had not been turned off. Furious about the incidents, Nanos suspended research at the laboratory and warned that he would fire "cowboys" who flouted the rules.

On 22 July, citing "almost suicidal denial" of security and safety practices, Nanos suspended 15 employees involved in the loss of the disks, along with four involved in the laser accident. All will continue to receive pay but are barred from entering the laboratory without an escort. The FBI is investigating the lost disks, and Nanos said some employees could face criminal charges.

The DOE-wide shutdown is affecting about a dozen laboratories that do classified work. None of the labs will be able to resume activity until they have performed a series of steps, including a complete inventory of portable disks, the creation of secure repositories for disks and other removable devices containing classified information, and a visit from an independent review team.

In the meantime, some researchers are becoming frustrated. In Los Alamos, for instance, residents report a growing number of cars sporting ironic bumper stickers that say "Striving for a Work-Free Safe Zone."

–DAVID MALAKOFF

Researchers say crossing avian and human flu viruses is crucial to understanding the threat of a new influenza pandemic, but they admit that they might create a monster

Tiptoeing Around Pandora's Box

Once again, the world is crossing its fingers. The avian influenza outbreak in Asia, already one of the worst animal-health disasters in history, has flared up in four countries; tens of thousands of birds are being killed in desperate attempts to halt the virus's spread. And again, the unnerving question arises: Could the outbreak of the H5N1 strain spiral into a human flu pandemic, a global cataclysm that could kill millions in a matter of months and shake societies to their core?

There is a way to find out, flu scientists say—but it's controversial.

Leaving nature to take its course, a pandemic could be ignited if avian and human influenza strains recombine—say, in the lungs of an Asian farmer infected with both—producing a brand-new hybrid no human is immune to. By mixing H5N1 and human flu viruses in the lab, scientists can find out how likely this is, and how dangerous a hybrid it would be.

Such experiments can give the world a better handle on the risks, but they could also create dangerous new viruses that would have to be destroyed or locked up forever in a scientific high-security prison. An accidental release—not so far-fetched a scenario given that the severe acute respiratory syndrome (SARS) virus managed to escape from three Asian labs in the past year—could lead to global disaster. Given their scientific merit, the World Health Organization (WHO) is enthusiastically promoting the experiments. But worried critics point out that there is no global mechanism to ensure that they are done safely.

Despite the concerns, such studies have already begun. In 2000, the U.S. Centers for Disease Control and Prevention (CDC) in Atlanta, Georgia, started experiments to create crossovers between the H5N1 strain isolated during a 1997 outbreak in Hong Kong and a human flu virus adapted for the lab. The study was suspended when CDC's flu researchers became overwhelmed by SARS and the new H5N1 outbreak, both in 2003, says CDC flu expert Nancy Cox, who led the work. But the agency plans to resume the work shortly with the H5N1 strain now raging in Asia.

Others are exploring the options as well. Virologist Albert Osterhaus of Erasmus University in Rotterdam, the Netherlands, is eager to try not just H5N1 but also other bird flu strains, such as H7N7. The Netherlands won't have the required high-level biosafety lab until late 2005, so Osterhaus is talking to

Risk assessment. The H5N1 influenza strain is highly lethal to humans, but whether it could trigger a pandemic is still uncertain.

researchers in France who do. In the United Kingdom, researchers at the Health Protection Agency, the National Institute for Biological Standards and Control, and universities are also discussing the idea. There are no concrete plans yet—in part because of a lack of funds—but there's a consensus that the studies are important and that Britain is well suited to do them, says influenza researcher Maria Zambon of the Health Protection Agency.

The aim of reassortment studies, as they're called, would not be to develop new countermeasures, says WHO's principal flu scientist, Klaus Stöhr, because researchers believe current drugs and an H5N1 vaccine in development would work against a pandemic strain as well. But the experiments would provide a badly needed way to assess the risk of a pandemic. If they indicate that a pandemic virus is just around the corner, health officials would further intensify their fight in Asia and go full-throttle in stashing vaccines and drugs; if not, they could breathe a little easier. "It's an extremely important question, and we have a responsibility to answer it," insists Stöhr.

The safety worries are legitimate, Stöhr concedes, and the work should be done only by labs with ample flu expertise and excellent safety systems—not the ones that let SARS out. "We don't want people just fiddling around," he says. He also downplays concerns that the results, when published, might help those who would unleash a pandemic on purpose. Anyone with the scientific smarts to do so can already find plenty of ideas in the literature, Stöhr asserts. Moreover, the studies are unlikely to produce anything that could not arise naturally, says Osterhaus: "You could create a monster. But it's a monster that nature could produce as well."

But critics beg to differ. "We've been debating whether to destroy the smallpox virus for years—and now we're planning to create something that's almost as dangerous?" asks Mark Wheelis, an arms-control researcher at the University of California, Davis. Wheelis also points out that there's no way to keep countries with poor safety records from getting in on the game. At the very least, there should be some global consensus on how to proceed, adds Elisa Harris, a researcher at the Center for International and Security Studies at the University of Maryland, College Park—although no formal mechanism for reaching it exists.

Mix and match
The H5N1 strain has been vicious to its human victims, killing 23 of 34 patients in Vietnam and Thailand this year. So far, however, every known patient had been in contact with infected birds; there's no evidence that the virus can jump from one person to the next—for now. But the virus could evolve inside one of its human hosts, acquir-

ing mutations that make it possible to infect humans directly, Stöhr says. Another scenario—one researchers believe sparked several previous influenza pandemics—is reassortment with a human flu virus in a person infected with both.

Influenza has a peculiar genome that's divided into eight loose segments, most of them containing precisely one gene. Each segment is copied separately in the host cell's nucleus; at the end of the reproduction cycle, all eight meet up with one another—and with envelope and membrane proteins—to form a new virus particle that buds from the host cell membrane to wreak havoc elsewhere. When a cell happens to be infected with two different strains, homologous segments can mix and match into new, chimeric viruses.

To create a worldwide outbreak, a newcomer must cause disease in humans and be transmissible between them, and its coat must look so new that no human immune system recognizes it. This is determined primarily by the two glycoproteins on the viral surface, hemagglutinin and neuraminidase—the "H" and "N" in names like H5N1. (Hemagglutinin comes in at least 16 different types, N in nine.) The current fear is that the Asian flu will keep its H5—which humans have never seen before—but swap enough of the remaining seven gene segments with those of a human strain to become more adept at replication in its new host.

During H5N1's first major outbreak in Hong Kong poultry in 1997, 18 people got sick and six died. But the outbreak was stamped out efficiently, and little was heard of H5N1 for 6 years—until it came roaring back last year. Given the magnitude of the current outbreak, the riddle is why reassortment has not yet taken place, says Stöhr. Reassortment studies could help explain whether the world has simply been lucky, or whether there's some barrier to reassortment of H5N1.

The experiments are straightforward. Researchers take a cell line such as MDCK or Vero cells, often used for virus isolation, and add both H5N1 and a currently circulating human strain, such as H3N2 or H1N1. Or they can use a slightly less natural technique called reverse genetics, with which virtually any combination of genes can be put into a flu virus. Any viable hybrid strains would be inoculated into mice; those that cause disease would move on to ferrets, a species very similar to humans in its susceptibility to influenza. Any strain that is pathogenic in

CREDITS: (TOP TO BOTTOM) C. S. GOLDSMITH AND J. M. KATZ/CDC; P. HUEY/SCIENCE
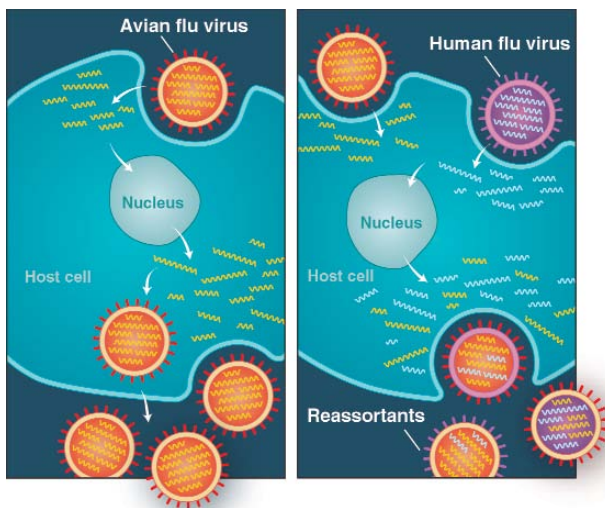
Two can tango. Flu virus genomes consist of eight segments, each of which is copied separately by the host cell (*left*). When two strains infect one cell, they can reassort (*right*).

ferrets and also jumps, say, from a sick animal to a healthy one in an adjacent cage could be humankind's next nightmare.

During its first round of experiments with the H5N1 strain, CDC managed to create several reassortants, Cox says, but it didn't get around to characterizing them; they're still sitting in a locked freezer in Atlanta.

Global risks, global review?
Most agree that such experiments are in a league of their own. Controversial flu studies were conducted in the past; for instance, researchers sequenced parts of the genome of the "Spanish flu" strain from 1918 (*Science*, 21 March 1997, p. 1793) and inserted its genes into other strains to find out why it was so deadly. But that didn't amount to a wholesale fishing expedition for pandemic strains. And because the 1918 strain was an H1 virus, just like one of the currently active ones, you'd expect at least some immunity to it in the human population, says Yoshihiro Kawaoka of the University of Tokyo and the University of Wisconsin, Madison, who studies the 1918 strain. With an H5 virus, in contrast, everyone would be vulnerable.

Yet although most countries have systems to review the safety and ethical aspects of run-of-the-mill scientific studies, none have formal panels to weigh studies that could, say, put the entire world at risk or be of potential help to bioterrorists. [The U.S. government has announced plans for a national biosecurity panel and a review system to fill that gap (*Science*, 12 March, p. 1595), but they have yet to be implemented.] So although CDC's first round of studies cleared

all the usual review hurdles at the agency, Cox says, nothing beyond that was considered necessary.

Since then, "the times have changed," Cox says. The H5N1 strain now plaguing Asia, with which CDC wants to work this time, appears to be more virulent than the 1997 version, and the specter of nefarious use of pathogens looms much larger. Moreover, the mishaps with SARS have made people jittery about labs' abilities to keep bugs on the inside. That's why Cox says she has consulted more extensively with colleagues inside and outside CDC, including experts such as Nobel laureate Joshua Lederberg and WHO. She also plans to seek approval from colleagues at the U.S. National Institutes of Health and the U.S. Food and Drug Administration.

But flu researcher Karl Nicholson of the University of Leicester, U.K., says there should be a more formal, global consensus on the necessity of the studies, who should conduct them, and how. For any country to undertake them on its own, he says, "is like a decision to start testing nuclear weapons unilaterally." WHO would be the best organization to start such a process, says Harris: The destruction of the smallpox virus has been debated at WHO, and an international panel there is overseeing experiments with it at CDC and in Russia.

But Stöhr believes existing safeguards suffice. The studies have been discussed widely with scientists in WHO's global flu lab network and at a recent flu meeting in Lisbon, he says, and have met with nothing but "overwhelming agreement." "If there are other voices, we will take them seriously," Stöhr adds—but for now, it's up to the labs to have their plans rigorously vetted by national authorities and get started.

Eventually, any strain with pandemic potential should be destroyed, he says. But there's no way to enforce this, and skeptics point out that the smallpox virus was slated for destruction, too—until the threat of bioterrorism created a movement to keep it alive, perhaps indefinitely, for defensive studies. In a way this discussion is moot, says Richard Webby of St. Jude Children's Research Hospital in Memphis, Tennessee. With flu strains readily available, anyone with a good knowledge of molecular biology could recreate a pandemic virus once it's discovered, he says. "You can destroy this virus," Webby says, "but it will never really be gone."
　　　　　　　　　　　　–MARTIN ENSERINK

Crime, Culpability, and the Adolescent Brain

This fall, the U.S. Supreme Court will consider whether capital crimes by teenagers under 18 should get the death sentence; the case for leniency is based in part on brain studies

When he was 17 years old, Christopher Simmons persuaded a younger friend to help him rob a woman, tie her up with electrical cable and duct tape, and throw her over a bridge. He was convicted of murder and sentenced to death by a Missouri court in 1994. In a whipsaw of legal proceedings, the Missouri Supreme Court set the sentence aside last year. Now 27, Simmons could again face execution: The state of Missouri has appealed to have the death penalty reinstated. The U.S. Supreme Court will hear the case in October, and its decision could well rest on neurobiology.

At issue is whether 16- and 17-year-olds who commit capital offenses can be executed or whether this would be cruel and unusual punishment, banned by the Constitution's eighth amendment. In a joint brief filed on 19 July, eight medical and mental health organizations including the American Medical Association cite a sheaf of developmental biology and behavioral literature to support their argument that adolescent brains have not reached their full adult potential. "Capacities relevant to criminal responsibility are still developing when you're 16 or 17 years old," says psychologist Laurence Steinberg of the American Psychological Association, which joined the brief supporting Simmons. Adds physician David Fassler, spokesperson for the American Psychiatric Association (APA) and the American Academy of Child and Adolescent Psychiatry, the argument "does not excuse violent criminal behavior, but it's an important factor for courts to consider" when wielding a punishment "as extreme and irreversible as death."

The Supreme Court has addressed some of these issues before. In 1988, it held that it was unconstitutional to execute convicts under 16, but it ruled in 1989 that states were within their rights to put 16- and 17-year-old criminals to death. Thirteen years later, it decided that mentally retarded people shouldn't be executed because they have a reduced capacity for "reasoning, judgment, and control of their impulses," even though they generally know right from wrong (see sidebar on p. 599). That is the standard Simmons's lawyers now want the court to extend to everyone under 18.

Test case. Christopher Simmons received the death penalty for a crime he committed at 17.

Cruel and unusual?

Simmons's lawyers argue that adolescents are not as morally culpable as adults and therefore should not be subject to the death penalty. They claim that this view reflects worldwide "changing standards of decency," a trend that has been recognized in many U.S. courts. Today, 31 states and the federal government have banned the juvenile death penalty. The latest to do so, Wyoming and South Dakota, considered brain development research in their decisions.

Putting a 17-year-old to death for capital crimes is cruel and unusual punishment, according to this reasoning. "What was cruel and unusual when the Constitution was written is different from today. We don't put people in stockades now," says Stephen Harper, a lawyer with the Juvenile Justice Center of the American Bar Association (ABA), which also signed an amicus curiae brief. "These standards mark the progress of a civilized society."

The defense is focusing on the "culpability of juveniles and whether their brains are as capable of impulse control, decision-making, and reasoning as adult brains are," says law professor Steven Drizin of Northwestern University in Chicago. And some brain researchers answer with a resounding "no." The brain's frontal lobe, which exercises restraint over impulsive behavior, "doesn't begin to mature until 17 years of age," says neuroscientist Ruben Gur of the University of Pennsylvania in Philadelphia. "The very part of the brain that is judged by the legal system process comes on board late."

But other researchers hesitate to apply scientists' opinions to settle moral and legal questions. Although brain research should probably take a part in policy debate, it's damaging to use science to support essentially moral stances, says neuroscientist Paul Thompson of the University of California, Los Angeles (UCLA).

Shades of gray

Structurally, the brain is still growing and maturing during adolescence, beginning its final push around 16 or 17, many brain-imaging researchers agree. Some say that growth maxes out at age 20. Others, such as Jay Giedd of the National Institute of Mental Health (NIMH) in Bethesda, Maryland, consider 25 the age at which brain maturation peaks.

Various types of brain scans and anatomic dissections show that as teens age, disordered-looking neuron cell bodies known as gray matter recede, and neuron projections covered in a protective fatty sheath, called white matter, take over. In 1999, Giedd and colleagues showed that just before puberty, children have a growth spurt of gray matter. This is followed by massive "pruning" in which about 1% of gray matter is pared down each year during the teen years, while the total volume of white matter ramps up. This process is thought to shape the brain's neural connections for adulthood, based on experience.

In arguing for leniency, Simmons's supporters cite some of the latest research that points to the immaturity of youthful brains, such as a May study of children and teens, led by NIMH's Nitin Gogtay. The team followed 13 individuals between the ages of 4 and 21, performing magnetic resonance imaging (MRI) every 2 years to track changes in the physical structure of brain tissue. As previous research had suggested, the frontal lobes matured last. Starting from the back of the head, "we see a wave of brain change moving forward into the front of the brain like a forest fire," says UCLA's Thompson, a co-author. The brain changes continued up to age 21, the oldest person they examined. "It's quite possible that the brain maturation peaks after age 21," he adds.

The images showed a rapid conversion

from gray to white matter. Thompson says that researchers debate whether teens are actually losing tissue when the gray matter disappears, trimming connections, or just coating gray matter with insulation. Imaging doesn't provide high enough resolution to distinguish among the possibilities, he notes: "Right now we can image chunks of millions of neurons, but we can't look at individual cells." A type of spectroscopy that picks out N-acetylaspartate, a chemical found only in neurons, shows promise in helping to settle the issue.

In addition to growing volume, brain studies document an increase in the organization of white matter during adolescence. The joint brief cites a 1999 study by Tomás Paus of McGill University in Montreal and colleagues that used structural MRI to show that neuronal tracts connecting different regions of the brain thickened as they were coated with a protective sheath of myelin during adolescence (*Science*, 19 March 1999, p. 1908).

In 2002, another study revealed that these tracts gained in directionality as well. Relying on diffusion tensor MRI, which follows the direction that water travels, Vincent Schmithorst of the Children's Hospital Medical Center in Cincinnati, Ohio, and colleagues watched the brain organize itself in 33 children and teens from age 5 to 18. During adolescence, the tracts funneled up from the spinal tract, through the brainstem, and into motor regions. Another linked the two major language areas. "The brain is getting more organized and dense with age," Schmithorst says.

Don't look at the light

Adults behave differently not just because they have different brain structures, according to Gur and others, but because they use the structures in a different way. A fully developed frontal lobe curbs impulses coming from other parts of the brain, Gur explains: "If you've been insulted, your emotional brain says, 'Kill,' but your frontal lobe says you're in the middle of a cocktail party, 'so let's respond with a cutting remark.'"

As it matures, the adolescent brain slowly reorganizes how it integrates information coming from the nether regions. Using functional MRI—which lights up sites in the brain that are active—combined with simple tests, neuroscientist Beatriz Luna of the University of Pittsburgh has found that the brain switches from relying heavily on local regions in childhood to more distributive and collaborative interactions among distant regions in adulthood.

One of the methods Luna uses to probe brain activity is the "antisaccade" test: a simplified model of real-life responses de-

Normal Brain Development

Age 5

Age 8

Age 12

Age 16

Age 20

>0.5
0.4
0.3
0.2
0.1
0.0

Gray Matter Volume

Maturing brain. An NIMH study of 13 individuals over a decade reveals a process—still under way in the late teens—in which gray matter is replaced throughout the cortex, starting at the rear.
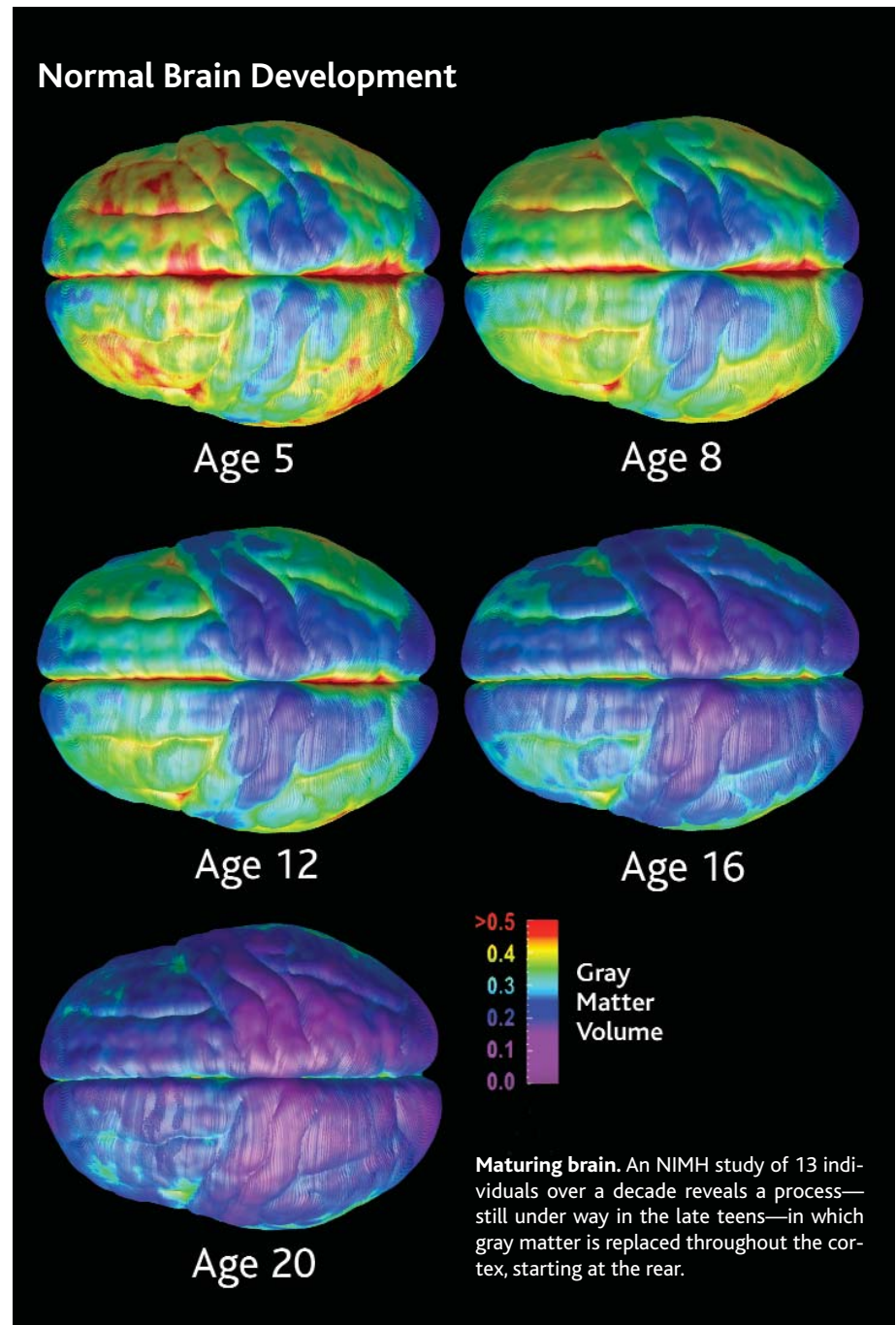
signed to determine how well the prefrontal cortex governs the more primitive parts of the brain. Subjects focus on a cross on a screen and are told that the cross will disappear and a light will show up. They are told not to look at the light, which is difficult because "the whole brainstem is wired to look at lights," says Luna.

Adolescents can prevent themselves from peeking at the light, but in doing so they rely on brain regions different from those adults use. In 2001, Luna and colleagues showed that adolescents' prefrontal cortices were considerably more active than adults' in this test. Adults also used areas in

the cerebellum important for timing and learning and brain regions that prepare for the task at hand.

These results support other evidence showing that teens' impulse control is not on a par with adults'. In work in press in *Child Development*, Luna found that volunteers aged 14 years and older perform just as well on the task as adults, but they rely mainly on the frontal lobe's prefrontal cortex, whereas adults exhibit a more complex response. "The adolescent is using slightly different brain mechanisms to achieve the goal," says Luna. Although the work is not cited in the brief, Luna says it

Adolescence: Akin to Mental Retardation?

The human brain took center stage in 2002 when the U.S. Supreme Court ruled against the death penalty for mentally retarded persons. In that case (*Atkins* v. *Virginia*), six of the nine justices agreed that executing a convict with limited intellectual capacity, Daryl Atkins, would amount to cruel and unusual punishment. Instructing the state of Virginia to forgo the death penalty in such cases, Justice John Paul Stevens wrote: "Because of their disabilities in areas of reasoning, judgment, and control of their impulses, [mentally retarded persons] do not act with the level of moral culpability that characterizes the most serious adult criminal conduct."

Last stop. In 2002, the Supreme Court rejected the death penalty (6–3) for mentally retarded persons.

When the case of Christopher Simmons, who committed murder at age 17, comes before the same justices in October, says law professor Steven Drizin of Northwestern University in Chicago, defense attorneys hope to equate juvenile culpability to that of mentally retarded persons. "Juveniles function very much like the mentally retarded. The biggest similarity is their cognitive deficit. [Teens] may be highly functioning, but that doesn't make them capable of making good decisions," he says. Brain and behavior research supports that contention, argues Drizin, who represents the Children and Family Justice Center at Northwestern on the amicus curiae brief for Simmons. The "standard of decency" today is that teens do not deserve the same extreme punishment as adults.

The Atkins decision provides advocates with a "template" for what factors should be laid out to determine "evolving standards of decency," says Drizin. These factors include the movement of state legislatures to raise the age limit for the death penalty to 18, jury verdicts of juvenile offenders, the international consensus on the issue, and public opinion polls. In 2002, the court also considered the opinions of professional organizations with pertinent knowledge, which is how the brain research comes into play. Last, the justices considered evidence that the mentally retarded may be more likely to falsely confess and be wrongly convicted—a problem that adolescents have as well.

–M.B.

clearly shows that "adolescents cannot be viewed at the same level as adults."

Processing fear

Other studies—based on the amygdala, a brain region that processes emotions, and research on risk awareness—indicate that teenagers are more prone to erratic behavior than adults. Abigail Baird and Deborah Yurgelun-Todd of Harvard Medical School in Boston and others asked teens in a 1999 study to identify the emotion they perceive in pictures of faces. As expected, functional MRI showed that in both adolescents and adults, the amygdala burst with activity when presented with a face showing fear. But the prefrontal cortex didn't blaze in teens as it did in adults, suggesting that emotional responses have little inhibition. In addition, the teens kept mistaking fearful expressions for anger or other emotions.

Baird, now at Dartmouth College in Hanover, New Hampshire, says that subsequent experiments showed that in teenagers the prefrontal cortex buzzes when they view expressions of people they know. Also, the children identified the correct emotion more than 95% of the time, an improvement of 20% over the previous work.

The key difference between the results, says Baird, is that adolescents pay attention to things that matter to them but have difficulty interpreting images that are unfamiliar or seem remote in time. Teens shown a disco-era picture in previous studies would say, "Oh, he's freaked out because he's stuck in the '70s," she says. Teens are painfully aware of emotions, she notes.

But teens are really bad at the kind of thinking that requires looking into the future to see the results of actions, a characteristic that feeds increased risk-taking. Baird suggests: Ask someone, "How would you like to get roller skates and skate down some really big steps?" Adults know what might happen at the bottom and would be wary. But teens

don't see things the same way, because "they have trouble generating hypotheses of what might happen," says Baird, partly because they don't have access to the many experiences that adults do. The ability to do so emerges between 15 and 18 years of age, she theorizes in an upcoming issue of the *Proceedings of the Royal Society of London*.

Luna points out that the tumultuous nature of adolescent brains is normal: "This transition in adolescence is not a disease or an impairment. It's an extremely adaptive way to make an adult." She speculates that risk-taking and lowered inhibitions provide "experiences to prune their brains."

With all the pruning, myelination, and reorganization, an adolescent's brain is unstable, but performing well on tests can make teens look more mature than they are. "Yes, adolescents can look like adults. But put stressors into a system that's already fragile, and it can easily revert to a less mature state," Luna says.

The amicus curiae brief endorsed by the APA and others also describes the fragility of adolescence—how teens are sensitive to peer pressure and can be compromised by a less-than-pristine childhood environment. Abuse can affect how normally brains develop. "Not surprisingly, every [juvenile offender on death row] has been abused or neglected as a kid," says ABA attorney Harper.

Biology and behavior

Although many researchers agree that the brain, especially the frontal lobe, continues to develop well into teenhood and beyond, many scientists hesitate to weigh in on the legal debate. Some, like Giedd, say the data "just aren't there" for them to confidently testify to the moral or legal culpability of adolescents in court. Neuroscientist Elizabeth Sowell of UCLA says that too little data exist to connect behavior to brain structure, and imaging is far from being diagnostic. "We couldn't do a scan on a kid and decide if they should be tried as an adult," she says.

Harper says the reason for bringing in "the scientific and medical world is not to persuade the court but to inform the court." Fassler, who staunchly opposes the juvenile death penalty, doesn't want to predict how the case will turn out. "It will be close. I'm hopeful that the court will carefully review the scientific data and will agree with the conclusion that adolescents function in fundamentally different ways than adults." And perhaps, advocates hope, toppling the death penalty with a scientific understanding of teenagers will spread to better ways of rehabilitating such youths.

–MARY BECKMAN

Mary Beckman is a writer in southeastern Idaho.

How Species Arise

Benjamin K. Blackman and Loren H. Rieseberg

One of the most important legacies of the modern synthesis is the articulation of the biological species concept (BSC). In his seminal 1942 work, *Systematics and the Origin of Species* (*1*), Ernst Mayr defined species as "groups of actually or potentially interbreeding natural populations, which are reproductively isolated from other such groups." Explicitly relating the definition of species to the process of speciation, the BSC has thrived—despite numerous hopeful alternatives—by inspiring a wealth of literature on reproductive isolation and gene flow. The last two decades in particular have brought major advances in molecular genetics, comparative analysis, mathematical theory, and molecular phylogenetics; speciation has consequently matured from a field fraught with untestable ideas to one reaching clear, well-supported conclusions.

> **Speciation**
> *by Jerry A. Coyne and H. Allen Orr*
>
> Sinauer, Sunderland, MA, 2004. 557 pp. $89.95. ISBN 0-87893-091-4. Paper, $54.95, £34.99. ISBN 0-87893-089-2.

Jerry Coyne and Allen Orr's *Speciation* provides a much-needed review of these developments. The exceedingly well-written and persuasive text eschews speculation. The authors instead resolutely develop testable criteria for distinguishing alternative hypotheses about evolutionary processes that may result in similar biological patterns, critically evaluate how theoretical and empirical results meet the burden of proof, and actively confront important caveats and unresolved questions with practical suggestions. It is a testament both to the authors and to the state of the field that the book provides such a robust picture of the origin of species.

The "species problem" that Coyne and Orr consider in the book is how do species arise. More specifically, why do sexually reproducing organisms fall into discrete clusters? With this question in mind, they choose a relaxed version of the BSC that allows for some gene flow among species as long as distinctiveness is maintained. Although Coyne and Orr recognize that the segregation of biological diversity into distinct clus-

The reviewers are in the Department of Biology, Indiana University, Bloomington, IN 47405, USA. E-mail: bkblackm@indiana.edu (B.K.B.) and lriesebe@indiana.edu (L.H.R.)

ters likely has multiple causes including ecology and history, the book concentrates almost exclusively on reproductive isolation. Given that the bulk of speciation research has focused on the origin of reproductive barriers, the authors' predominant focus is, if nothing else, appropriate for their chosen task. And although the book's subject matter and neontological approach may prove unhelpful to systematists or paleontologists, such readers may find some appeasement in the favorable treatment of species selection and the thorough yet concise appendix that discusses the relative merits and pitfalls of several major species concepts.

Proceeding from their premise that studying speciation is largely synonymous with studying reproductive isolation, Coyne and Orr explore what we know about where, when, and how isolating barriers evolve. Following Mayr, they argue that speciation most often occurs where populations are geographically isolated or "allopatric." The

Evidence for allopatry. These three congeneric Pacific wrasse (top to bottom: *Halichoeres trimaculatus*, *H. margaritaceus*, and *H. hortulanus*) were painted for David Starr Jordan and Alvin Seale's *The Fishes of Samoa* (*5*) by the Japanese artist Kako Morita. (Morita's work was published with the help of Theodore Roosevelt, who interceded after a government committee ruled the plates were too expensive to print.) Jordan argued that geographical barriers were required for speciation.

broad range of theoretical conditions under which reproductive isolation evolves in allopatry, experimental evolution of reproductive barriers in isolated laboratory populations, and abundant examples of speciation events associated with vicariance events or isolation on islands all strongly support this position. However, unlike Mayr, Coyne and Orr reach a more favorable though still unenthusiastic view of sympatric speciation, one largely based on the development of theoretical models with increasingly realistic assumptions that indicate sympatric speciation could occur. They find empirical data to be less compelling: only three case studies not involving polyploid or hybrid speciation meet their criteria for a biogeographic and evolutionary history that makes an allopatric phase highly unlikely.

An examination of when and how isolating barriers evolve forms the core of *Speciation*. Far from being a dry catalog of mechanisms and well-known examples, these chapters offer engaging discussions that aim to sharpen how we define, detect, and measure isolating barriers; challenge us to decipher the evolutionary rather than current importance of these barriers; and synthesize evidence regarding their genetics and evolution. The treatment of mechanical isolation is an often-entertaining read, and the overall attempt to outline and then disentangle how natural and sexual selection may act to promote nonecological or behavioral forms of isolation is methodical and enlightening.

Speciation convincingly presents evidence for several once-unpopular theories that have returned to dominate current thinking. Most important among these is the primacy of natural and sexual selection over drift in driving speciation. Signatures of positive selection on genes involved in postzygotic isolation and reproductive proteins as well as experimental evidence from both the lab and field connect adaptation and sexual selection to reproductive isolation. Another major finding is the congruence of the Dobzhansky-Muller model for the evolution of postzygotic isolation with the genetics of hybrid incompatibilities in many natural systems. In contrast, classical models of chromosomal speciation remain unpopular. Instead, chromosomal rearrangements are now cast as facilitators, rather than causal agents, of reproductive isolation because reduced recombination within these regions restricts gene flow, thereby enabling the accumulation of selected differences and hybrid incompatibilities.

The authors take cautious views on controversial questions like reinforcement, sympatric speciation, and diploid hybrid (or "recombinational") speciation. For al-

though recent theoretical advances demonstrate each phenomenon can occur under a nontrivial set of conditions, conclusive empirical evidence that they occur in nature only exists for the third process. Even so, the authors believe comparative analyses and further case studies will prove fruitful avenues for determining if and how often these processes operate in nature.

Coyne and Orr, who are *Drosophila* population geneticists at the University of Chicago and the University of Rochester, respectively, provide remarkably lucid explanations of speciation phenomena in other groups of organisms, alleviating prepublication fears that the book would be dominated by flies. Their discussion of polyploidy, for example, is perhaps the best review of a predominantly botanical literature by zoologists. Treatments of other plant-related topics like mating system isolation or hybridization are insightful as well, but may raise eyebrows. For instance, unlike most botanical discussions of mating system evolution (*2*), Coyne and Orr argue that the shift from outbreeding to selfing is not a kind of reproductive isolation because gene flow is reduced as much within as among taxa. Likewise, botanists may find an otherwise excellent treatment of recombinational speciation to be tilted toward the evolution of postzygotic barriers through hybridization as opposed to the contribution of new hybrid gene combinations to ecological differentiation and species establishment. Lastly, the authors downplay increasingly widespread phylogenetic evidence of cryptic introgression or hybrid speciation in the plant, and now even animal, literature.

The book is a rich and thorough review, critique, and synthesis of recent literature that is sure to become a classic read for anyone interested in speciation. As the authors' purpose is to reflect on the value of various approaches to evolutionary questions and point out areas ripe for further investigation, *Speciation* is not a textbook that pauses to give broad introductions; many methods and terms are referred to in passing well before being defined in later chapters. Despite this, Coyne and Orr's descriptions and logical evaluations of theoretical and empirical work are remarkably clear and straightforward, a considerable achievement because the book covers material from complicated mathematics to rigorous molecular genetics. An excellent book for a graduate seminar, *Speciation* should also be interesting and accessible to scientists from diverse backgrounds.

Notably, many important results that support Coyne and Orr's conclusions in the book have only been published in the last year. For instance, two of the four genes known to underlie hybrid incompatibilities were identified only recently, and their analysis adds great support to the role of selection over drift in the evolution of these barriers (*3*, *4*). With such research ongoing, and now with *Speciation* as a guide, the authors' wish that their book "will stimulate younger scientists to pursue their own work on speciation" will certainly be fulfilled.

References
1. E. Mayr, *Systematics and the Origin of Species* (Columbia Univ. Press, NY, 1942).
2. D. A. Levin, *Evol. Biol.* **11**, 185 (1978).
3. D. A. Barbash, D. F. Siino, A. M. Tarone, J. Roote, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 5302 (2003).
4. D. C. Presgraves, L. Balagopalan, S. M. Abmayr, H. A. Orr, *Nature* **423**, 715 (2003).
5. D. S. Jordan, A. Seale, *Bull. U.S. Bur. Fish.* **25**, 173 (1906).

EVOLUTION

Hunting for Origins

R. Andrew Cameron

The term "Cambrian explosion" is really a metaphor because the phenomenon named here is neither an explosion nor did it happen in the Cambrian. Yes, there appeared in Cambrian rocks (Chengjiang formation) dated 520 million years ago representatives of almost all major groups of animals. But newly estimated rates of change in protein and DNA sequences calibrated to well-dated fossils set the divergences of these major groups to a time well before the Cambrian (*1*). Given this apparent contradiction, many who study animal evolution reckon that the early animals in these lineages were small and soft-bodied, resulting in a poor fossil record. Perhaps the conditions of the Cambrian environment allowed the rapid appearance of hard skeletal parts, greatly favored fossilization, or both.

> **On the Origin of Phyla**
> *by James W. Valentine*
>
> University of Chicago Press, Chicago, 2004. 638 pp. $55, £38.50. ISBN 0-226-84548-6.

In this context James Valentine (an emeritus professor of integrative biology at the University of California, Berkeley) delivers a new book aimed at explaining the origin of the highest taxonomic groups of metazoans, *On the Origin of Phyla*. Considering the great variety of existing animals and the explanations for their elaboration, this is no easy job. There has been a steady trickle of books, some best sellers, offered to incorporate Darwinian evolution into a synthesis explaining the origin of higher taxa, but none have come to represent the field the way that

The reviewer is in the Division of Biology, Mail Code 139-74, 1200 East California Boulevard, California Institute of Technology, Pasadena, CA 91125, USA. E-mail: acameron@caltech.edu

An odd ecdysozoan. Phylogenetic analyses using morphological traits place *Opabinia* and the anomalocarids in a clade basal to Arthropoda or Onychophora. But an alternative interpretation of the imputed lobopodial structures in this predator from the Burgess Shale suggests it is a stem-group arthropod. (*6*)

Darwin's *The Origin of Species* has. Perhaps this is due in part to the modern fragmentation of studies in biology and geology into specialized areas, which must be integrated to build an explanation, for as Valentine points out, ecologists have sought environmental explanations, developmental biologists have offered mechanisms for achieving variety of form, and so on. But there has recently been a convergence of new data from several areas of research that is especially relevant to these questions. Witness the new finds of fossil pre-Cambrian embryos from the Doushantuo Formation of southwest China, estimated to be 40 to 55 million years older than the base of the Cambrian (*2*); the maturing consensus emerging from molecular phylogenetics; and advances in the comparative molecular biology of development. Valentine has organized the book into sections that reflect these converging areas of study.

The question of when and how higher taxonomic groups like phyla evolved differs markedly from the one Darwin addressed 145 years ago in *The Origin of Species*. It is not simply different in scale but also in quality. Although it is somewhat easier to see how changes in single genes can lead to differences among species that render some more capable of surviving in particular environments, it is more difficult to account for the many changes that lead to entirely different bodyplans as a simple accumulation of single-gene effects. For example, marine stickleback fishes possess bony plates and spines that presumably prevent predation, while their freshwater relatives show a loss of this armor through changes that can be attributed to a single gene (*3*, *4*). However, en-tire organ systems or embryonic germ layers, features that distinguish higher taxa, can be explained in terms of the gene regulatory networks whose architecture is hardwired into the genome. The simplest of these networks leads to the specification of differentiated cell types, something like the cell morphotypes that Valentine defines.

Networks that underlie morphogenetic pattern formation programs defining clade-specific body parts and bodyplans are more complex (*5*). These networks are assembled from cis-regulatory interactions that operate at the genome level. They represent the heritable process by which the genome specifies the organism. From the structure of these networks emerge obvious ways that major changes can happen through changes in the links among regulatory genes like transcription factors. For instance, a morphogenetic program may evolve with relatively minimal changes to establish a new spatial domain of expression for a cell-differentiation program, and the resultant animal has a new body part. Although Valentine skirts around this mechanistic model for the evolution of developmental programs in his definitions of the hierarchy of genomes, genes, and their possible sources of change, he does not incorporate a molecular model in his final synthesis.

The intellectually greedy might flip to the back of the book, skipping over the sections on the evidence for the origins of metazoan phyla and the descriptions of the phyla themselves, even though these compilations (particularly the section on the fossil record) are the work's strong points. In the third section, Valentine paints his view of the evolution of the phyla, with an emphasis on the work of morphologists and paleontologists (a portion of which he contributed). He fleshes out a scenario of variation and extinction that unites the detailed descriptions of the first two sections. The pictures he delineates here reveal correlations uniting different levels of biological organization, but absent are firm statements about causal mechanisms from which predictions could be made. Only in the section's closing pages are genomes considered.

In view of the volatility of the ideas and the controversy that still exist in this particular area of evolutionary biology, one might argue that it is too early to explain the causes of the origin of phyla. But as Valentine aptly points out, the time will never be exactly right: there are always more information to incorporate and more ideas to organize. Though too heavy with data to be carried in the kit bag, *On the Origin of Phyla* is a likely candidate for the bookshelves of those who hunt for pre-Cambrian fossils or the historical patterns in DNA sequences.

References and Notes

1. K. J. Peterson *et al., Proc. Natl. Acad. Sci. U.S.A.* **101**, 6536 (2004).
2. J.-Y. Chen *et al., Science* **305**, 218 (2004); published online 3 June 2004 (10.1126/science.1099213).
3. W. A. Cresko *et al., Proc. Natl. Acad. Sci. U.S.A.* **101**, 6050 (2004).
4. J. S. McKinnon *et al., Nature* **429**, 294 (2004).
5. E. H. Davidson *et al., Science* **295**, 1669 (2002).
6. This reconstruction is from the Smithsonian Institution's Traveling Exhibition *The Burgess Shale: Evolution's Big Bang*, at the Sternberg Museum of Natural History, Hays, KS, through 24 October; the Burke Museum of Natural History and Culture, Seattle, WA, 13 November 2004 to 1 May 2005; and the Sam Noble Oklahoma Museum of Natural History, Norman, OK, 21 May to 27 November 2005.

POLICY FORUM

The Bush Administration's Approach to Climate Change

Spencer Abraham

As a signatory to the United Nations Framework Convention on Climate Change (UNFCCC), the United States shares with many countries its ultimate objective: stabilization of greenhouse gas concentrations in the atmosphere at a level that prevents dangerous interference with the climate system. Meeting this UNFCCC objective will require a long-term commitment and international collaboration.

President Bush's policy on climate change harnesses the power of markets and technological innovation, maintains economic growth, and encourages global participation. Although climate change is a complex and long-term challenge, the Bush administration recognizes that there are cost-effective steps we can take now.

Near-Term Policies and Measures

In 2002, President Bush set a national goal to reduce the greenhouse gas intensity (1) of the U.S. economy by 18% by 2012. This goal sets America on a path to slow the growth in greenhouse gas emissions and— as the science justifies and the technology allows—to stop and reverse that growth as needed to meet the UNFCCC goal (2). Our approach focuses on reducing emissions while sustaining the economic growth needed to finance investment in new, clean energy technologies. The administration estimates that this commitment will achieve about 100 million metric tons of carbon equivalent (MMTCe) of reduced emissions in 2012, with more than 500 MMTCe in cumulative savings over the decade (3).

To this end, the administration has developed an array of policy measures, including financial incentives and voluntary programs. For example, our Climate VISION (4), Climate Leaders (5), and SmartWay Transport Partnership (6) programs work with industry for voluntary reduction of emissions. The Department of Agriculture is using its conservation programs to provide an incentive for actions that increase carbon

The author is the U.S. Secretary of Energy, 1000 Independence Avenue, SW, Washington, DC 20585, USA.

sequestration (7). We also are pursuing many energy supply technologies with comparatively low or zero CO_2 emissions profiles, such as solar, wind, geothermal, bioenergy, and combined heat and power. The president has proposed more than $4 billion in tax credits as incentives for these and other energy-efficient technologies over the next 5 years (3). Last year, the Bush administration increased fuel economy standards for new light trucks and sport utility vehicles by 1.5 miles per gallon over the next three model years, leading to the estimated avoidance of 9.4 MMTCe of emissions (8).

While acting to slow the pace of greenhouse gas emissions in the near term, the United States is laying a strong scientific and technological foundation to reduce uncertainties, to clarify risks and benefits, and to develop realistic mitigation options to meet the UNFCCC objective.

Advancing Climate Change Science

In 2001, President Bush commissioned the National Research Council (NRC) to examine the state of our knowledge and understanding of climate change science. The NRC's report (9) makes clear that there are still important gaps in our ability to measure the impacts of greenhouse gases on the climate system. Major advances in understanding and modeling of the factors that influence atmospheric concentrations of greenhouse gases and aerosols, as well as the feedbacks that govern climate sensitivity, are needed to predict future climate change with greater confidence.

Last summer, the Climate Change Science Program (CCSP) released a new strategic plan that addresses these gaps (10). The plan is organized around five goals: (i) improving our knowledge of climate history and variability; (ii) improving our ability to quantify factors that affect climate; (iii) reducing uncertainty in climate projections; (iv) improving our understanding of the sensitivity and adaptability of ecosystems and human systems to climate change; and (v) exploring options to manage risks. Annually, almost $2 billion is spent on climate change science by the federal government.

A review of the CCSP plan by NRC shows the administration is on the right track. While concern was expressed about future funding to execute the plan, the NRC concluded that it "articulates a guiding vision, is appropriately ambitious, and is broad in scope" (11).

NRC's report also identified the real need for a broad global observation system to support measurements of climate variables. Last June, the United States hosted more than 30 nations at the inaugural Earth Observation Summit, out of which came a commitment to establish an intergovernmental, comprehensive, coordinated, and sustained Earth observation system. The data collected by the system will be used to create better climate models, to improve our knowledge of the behavior of CO_2 and aerosols in the atmosphere, and to develop strategies for carbon sequestration.

Accelerating Climate Change Technology Development

The Bush administration also is moving ahead on advanced technology options that have the potential to substantially reduce, avoid, or sequester future greenhouse gas emissions. About 80% of current greenhouse gas emissions are energy related, and, although projections vary considerably, a tripling of energy demand by 2100 is not unimaginable (12). Therefore, to provide the energy necessary for continued economic growth while we reduce greenhouse gas emissions, we may have to develop and deploy cost-effective technologies that alter the way we produce and use energy.

By 2100, more than half of the world's energy may have to come from low- or zero-emission technologies to attain the UNFCCC goal (13). The pace and scope of needed change will be driven partially by future trends in greenhouse gas emissions that, like climate sensitivity, are uncertain. The complex relations among population growth; economic development; energy demand, mix, and intensity; resource availability; technology; and other variables make it impossible to accurately predict future greenhouse gas emissions on a 100-year time scale.

The Climate Change Technology Program (CCTP) was created to coordinate and prioritize the federal government's nearly $3 billion annual investment in climate-related technology research, development, demonstration, and deployment (RDD&D). Using various analytical tools, CCTP is assessing different technology options and their potential contributions to

reducing greenhouse gas emissions. Given the tremendous capital investment in existing energy systems, the desired transformation of our global energy system may take decades or more to implement fully. A robust RDD&D effort can make advanced technologies available sooner rather than later and can accelerate modernization of capital stock at lower cost and with greater flexibility.

CCTP's strategic vision has six complementary goals: (i) reducing emissions from energy use and infrastructure; (ii) reducing emissions from energy supply; (iii) capturing and sequestering CO_2; (iv) reducing emissions of other greenhouse gases; (v) measuring and monitoring emissions; and (vi) bolstering the contributions of basic science (*14*).

Ten federal agencies support a portfolio of activities within this framework. Annually, more than $700 million is being spent to advance energy efficiency technologies (plus $500 million for accelerated deployment), and more than $200 million supports renewable energy. Many activities build on existing work, but the Bush administration also has expanded and realigned some activities and launched new initiatives in key technology areas to support the CCTP's goals.

In his 2003 State of the Union address, President Bush made a commitment to the development of a hydrogen economy, pledging $1.7 billion over 5 years for his Hydrogen Fuel Initiative and Freedom-CAR Partnership to develop hydrogen fuel cell–powered vehicles. The transition to hydrogen as a major energy carrier over the next few decades could transform the nation's energy system and create opportunities to increase energy security by making better use of diverse domestic energy sources for hydrogen production and to reduce emissions of air pollutants and CO_2 (*15*). Where hydrogen is produced from fossil fuels, we must also address carbon capture and sequestration.

To help coordinate and leverage ongoing work overseas, the United States led the effort to form the International Partnership for the Hydrogen Economy (IPHE). IPHE will address the technological, financial, and institutional barriers to hydrogen and will develop internationally recognized standards to speed market penetration of the new technologies.

The administration also is pursuing next-generation nuclear energy as a zero-

Potential contributions to emissions reduction
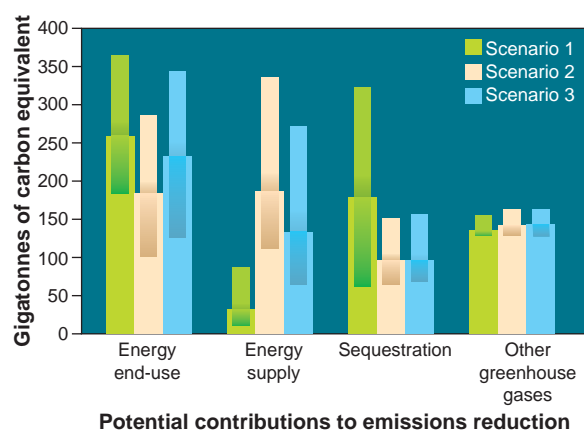
Potential ranges of greenhouse gas emissions reductions to 2100 by category of activity for three technology scenarios characterized by viable carbon sequestration (scenario 1); dramatically expanded nuclear and renewable energy (scenario 2); and novel and advanced technologies (scenario 3) (*14*).

emissions energy supply choice. The Generation IV International Forum, with nine other nations as partners, is working on reactor designs that are safe, economical, secure, and able to produce new products, such as hydrogen. Six promising technologies have been selected as candidates for future designs and could be ready as early as 2015. In 2003, President Bush announced that the United States would join the ITER project to develop fusion as an energy source. Although the technical hurdles are substantial, the promise of fusion is simply too great to ignore.

Carbon capture and sequestration is a central element of CCTP's strategy because for the foreseeable future, fossil fuels will continue to be the world's most reliable and lowest-cost form of energy. It is unrealistic to expect countries—particularly developing countries—with large fossil reserves to forgo their use. A realistic approach is to find ways to capture and store the CO_2 produced when these fuels are used.

The Department of Energy is currently working on 65 carbon sequestration projects around the country. In the last 2 years, we have increased the budget for these activities 23% to $49 million. The multilateral Carbon Sequestration Leadership Forum, a presidential initiative inaugurated in June 2003 with 16 partners, will set a framework for international collaboration on sequestration technologies.

The forum's partners are eligible to participate in FutureGen, a 10-year, $1 billion government-industry effort to design, build, and operate the world's first emissions-free coal-fired power plant. This project, which cuts across many CCTP strategic goals, will employ the latest technologies to generate electricity, produce

hydrogen, and sequester CO_2 from coal. Through this research, clean coal can remain part of a diverse, secure energy portfolio well into the future.

These initiatives and other technologies in the CCTP portfolio (*16*) could revolutionize energy systems and put us on a path to ensuring access to clean, affordable energy supplies while dramatically reducing greenhouse gas emissions. The figure, left, offers a glimpse of the range of emissions reductions new technologies might make possible in energy end use, energy supply, carbon sequestration, and other greenhouse gases on a 100-year scale and across a range of uncertainties.

The Bush administration has developed a comprehensive strategy on climate change that is informed by science, emphasizes innovation and technological solutions, and promotes international collaboration to support the UNFCCC objective. Although the scientific and technology challenges are considerable, the president remains committed to leading the way on climate change at home and around the world.

References and Notes

1. Measured as the ratio of greenhouse gases (carbon equivalent) emitted per real gross domestic product.
2. See www.whitehouse.gov/news/releases/2002/02/addendum.pdf.
3. *Global Climate Change Policy Book: A New Approach* (The White House, Washington, DC, 14 February 2002); available at www.whitehouse.gov/news/releases/2002/02/climatechange.html.
4. See www.climatevision.gov.
5. See www.epa.gov/climateleaders.
6. See www.epa.gov/smartway.
7. See www.usda.gov/news/releases/2003/06/fs-0194.htm.
8. National Highway Traffic Safety Administration, *Final Environmental Assessment: National Highway Traffic Safety Administration Corporate Average Fuel Economy (CAFE) Standards* (NHTSA, Washington, DC, 2003); available at: www.nhtsa.dot.gov/cars/rules/cafe/docs/239533_web.pdf.
9. National Research Council, *Climate Change Science: An Analysis of Some Key Questions*, Committee on the Science of Climate Change (National Academy Press, Washington, DC, 2001), pp. 20–21.
10. CCSP, *Strategic Plan for the U.S. Climate Change Science Program* (CCSP, Washington, DC, July 2003); available at www.climatescience.gov.
11. National Research Council, *Implementing Climate and Global Change Research: A Review of the Final U.S. Climate Change Science Program Strategic Plan* (National Academies Press, Washington, DC, 2004), p. 1.
12. Intergovernmental Panel on Climate Change, "An overview of the scenario literature," *Emissions Scenarios* (Cambridge Univ. Press, Cambridge, 2000).
13. See, for example, K. Caldeira, A. K. Jain, M. I. Hoffert, *Science* **299**, 2052 (2003).
14. CCTP, *U.S. Climate Change Technology Program Draft Strategic Plan: Vision and Framework* (CCTP, Washington, DC, in preparation); see www.climatetechnology.gov.
15. National Research Council, *The Hydrogen Economy: Opportunities, Costs, Barriers, and R&D Needs* (National Academies, Washington, DC, 2004).
16. CCTP, *Research and Current Activities* (CCTP, Washington, DC, 2003); available at www.climatetechnology.gov.

PERSPECTIVES

Ecogenomics Benefits Community Ecology

Marcel Dicke, Joop J. A. van Loon, Peter W. de Jong

Understanding how individual traits of organisms affect both their interactions with other species and the dynamics of the ecosystem community is an important challenge confronting evolutionary ecologists. A major problem has been the lack of genotypes that differ exclusively in the trait of interest as well as a lack of techniques for assessing expression of such genotypes. Recent breakthroughs in molecular biology have provided ecologists with exciting new tools with which to pursue an ecogenomics strategy (*1*, *2*). Now ecologists can perform delicate genetic manipulations to obtain well-characterized genotypes, which, together with mechanistic knowledge of phenotypic plasticity, provide information on the effects of individual traits on species interactions in communities. A major obstacle to the rapid incorporation of molecular tools into ecological studies, however, is that most model organisms used by molecular biologists do not yet match the model organisms favored by ecologists. This is changing, as exemplified by the study of Kessler *et al.* (*3*) published on page 665 of this issue. These authors incorporate molecular techniques into an ecological field study to investigate the effects of three genes on the interactions between native tobacco plants and their natural insect herbivores.

For more than a decade, Baldwin and colleagues have analyzed the responses of native tobacco plants induced by insect herbivores, with methods ranging from transcriptome analyses in the laboratory (*1*) to field studies of phenotypically manipulated plants (*4*). They have developed three tobacco plant lines in which they have genetically silenced one of three genes (encoding lipoxygenase, hydroperoxide lyase, or allene oxide synthase) of the oxylipin signaling pathway. Oxylipin

Enhanced online at www.sciencemag.org/cgi/content/full/305/5684/618

signaling mediates both direct (toxins) and indirect (natural enemies) plant defenses against herbivorous attack. Exposing these three tobacco plant genotypes together with a wild-type control to a natural insect community in Utah enabled the investigators to elucidate the effects of the three genes on host-plant selection by insect herbivores (*3*).

Plant-insect interactions comprise a major biological interface in terrestrial ecosystems (*5*) (see the Perspective on page 619). The ecology of these interactions has been the subject of intensive studies not only because such interactions are of fundamental interest, but also because they provide the basis of ecologically sound management of agricultural pests. Two major investigative themes are the complex of direct and indirect defenses that plants employ against insects, and the adaptations of insects to these

plant defenses (*5*). Plant defenses may affect the performance of herbivorous insects directly, or they may do so indirectly by enhancing the effectiveness of the herbivores' natural enemies (see the figure). Both types of defense may be either constitutive (switched on all the time) or may be induced only in response to herbivore attack (*6*).

In the new work, Kessler *et al.* (*3*) investigate the phenotypic effects of silencing genes known to be involved in direct and indirect defenses against insect herbivores. Under field conditions and after silencing of lipoxygenase-dependent signaling, the authors reveal that the tobacco plants not only were more susceptible to herbivory, but also showed a reduction in their emission of plant volatile organic compounds (which may affect attraction of the herbivore's enemies) in response to herbivore attack. The up-regulation of defense-related genes and the down-regulation of photosynthesis-related genes was observed in the wild-type control plants but not in the engineered plants. Kessler and Baldwin (*4*) previously demonstrated that induced tobacco volatiles attract predators that remove herbivore eggs. Given that silencing of the lipoxygenase pathway resulted in a reduction of volatile emission

Beating back the bugs. A tritrophic system consisting of plants, insect herbivores, and their natural enemies. This particular system comprises cabbage plants (*Brassica oleracea*) (**B**), herbivorous larvae of the cabbage white butterfly (*Pieris brassicae*) (**A**), and parasitoids, *Cotesia glomerata* (**C**), that attack *P. brassicae* caterpillars (C). Damage caused by caterpillars feeding on cabbage plants up-regulates the expression of various genes in the plants, which are visualized as red spots in the microarray (**E**), and down-regulates the expression of other genes (green spots). Herbivory induces up-regulation of the biosynthesis of certain types of glucosinolates (**D**), toxic secondary metabolites characteristic of the Brassicaceae that mediate a direct defense against herbivorous insects. Additionally, the emission of dozens of volatile organic compounds, each represented by a peak in the gas chromatogram (**F**), is induced by herbivory. These herbivore-induced volatiles act as an indirect defense by attracting parasitoids that lay eggs in the caterpillars. Shown are the green-leaf volatile (*Z*)-3-hexen-1-ol and the terpenoid 1,8-cineole, representatives of two dominant classes of volatiles emitted by cabbage (F).

The authors are in the Laboratory of Entomology, Wageningen University, Post Office Box 8031, NL-6700 EH Wageningen, Netherlands. E-mail: marcel.dicke@wur.nl

(*3*), a logical follow-up experiment would be to assess the effects of gene silencing on plant consumers at higher trophic levels.

Other studies have exploited transgenic plants in the laboratory to investigate the effects of individual genes on ecological interactions, such as the costs of resistance to a pathogen (direct defense) (*7*) or the attraction of the enemies of herbivorous insects (indirect defense) (*8*). Such laboratory studies are valuable for assessing the effects of genes on particular interactions. However, the ultimate assessment of a gene's function should be made under field conditions, and it is here that Kessler *et al*. (*3*) make their biggest contribution to the study of direct plant defense against herbivory.

Plants are endowed with a remarkable capacity to compensate for herbivore damage (*9*); the relationship between plant damage and seed production is complex and context dependent. The selective advantage of induced plant defenses should be detectable at the level of plant fitness, the ultimate evolutionary currency (see the Perspective on page 619). However, studies of transgenic plants in the field are subject to regulatory constraints—particularly with regard to the dissemination of transgenic plant pollen—which limits progress in this area.

Apart from exploiting the new opportunities that transgenic plants offer, ecologists can take advantage of other molecular tools to make unprecedented progress in understanding the ecology of plant-insect interactions. Adaptations of insects to plant defenses may be spatially variable and may lead to local variations in the genetics underlying such adaptations (*10*). Consequently, investigating the interaction between selective regimes on the one hand and migration, gene flow, and demography on the other can answer important evolutionary questions—for example, how to explain the present distribution of an adaptation to a plant defense? This question can be analyzed by integrating genetic data at neutral loci and loci under selection—an approach called population genomics (*11*)—with ecological data and spatial modeling. The population genomics approach involves genome-wide sampling of markers and comparing the variation in putative neutral markers with that in markers thought to be associated with genomic regions under selection. This enables the distinction between genome-wide effects (for example, those caused by migration) and locus-specific effects (for example, those influenced by selection). Such an approach will lead to new insights into the spreading of adaptive traits and the geographic distribution of these traits.

We have highlighted only two of the many molecular strategies that can be adopted by evolutionary ecologists. By adopting such molecular strategies, ecologists can study mechanism and function integratively as both the expressed genotype and the performance of the phenotype with other members of the community are known. Molecular techniques provide evolutionary ecologists with many more opportunities. As a representation of an organism's phenotype, microarrays may be used to assess phenotypic variation under natural conditions or after artificial selection. Molecular expression markers that have been selected on the basis of genomic information may be used to screen segregating populations for genotypes that show quantitative variation in the expression of a gene of interest (*12*) rather than the qualitative difference represented by a silenced individual gene and its wild-type control. This provides an even better set of genotypes for addressing questions about the evolutionary ecology of plant-insect interactions. The important contribution of Kessler *et al*. (*3*) will likely mark the beginning of a big leap forward in our comprehension of the ecology of plant-insect interactions.

References and Notes
1. A. Kessler, I. T. Baldwin, *Annu. Rev. Plant Biol.* **53**, 299 (2002).
2. M. Dicke, R. M. P. van Poecke, J. G. de Boer, *Basic Appl. Ecol.* **4**, 27 (2003).
3. A. Kessler, R. Halitschke, I. T. Baldwin, *Science* **305**, 665 (2004); published online 1 July 2004 (10.1126/science.1096931).
4. A. Kessler, I. T. Baldwin, *Science* **291**, 2141 (2001).
5. L. M. Schoonhoven, T. Jermy, J. J. A. van Loon, in *Insect-Plant Biology. From Physiology to Evolution* (Chapman & Hall, London, 1998).
6. M. Dicke, J. J. A. van Loon, *Entomol. Exp. Appl.* **97**, 237 (2000).
7. J. Bergelson, C. B. Purrington, C. J. Palm, J. C. Lopez-Gutierrez, *Proc. R. Soc. London B* **263**, 1659 (1996).
8. R. M. P. van Poecke, M. Dicke, *J. Exp. Bot.* **53**, 1793 (2002).
9. S. Y. Strauss, A. A. Agrawal, *Trends Ecol. Evol.* **14**, 179 (1999).
10. J. N. Thompson, *Am. Nat.* **153**, S1 (1999).
11. W. C. Black, C. F. Baer, M. F. Antolin, N. M. DuTeau, *Annu. Rev. Entomol.* **46**, 441 (2001).
12. E. E. Schadt *et al*., *Nature* **422**, 297 (2003).
13. The authors are supported by a VICI grant from the Netherlands Organisation for Scientific Research.

ECOLOGY

Herbivores Rule

Robert J. Marquis

Ever since Bates, Darwin, and Wallace, ecologists and evolutionary biologists have strived to uncover the causes of high species diversity in tropical regions. In the case of plant diversity, early studies in the tropical regions of Central and South America emphasized the contribution of biotic interactions, particularly the effects of plant-eating predators (herbivores) on plant survival. In contrast, efforts in tropical Asia have centered on abiotic factors, such as soil nutrients and water, describing how the species composition of forests shifts with changes in soil type. It is likely that neither factor works alone. Tropical forests grow on a patchwork of different soil types such that many tree and plant species are uniquely adapted to particular soil (edaphic) conditions. For example, on tropical white-sand soils—which are extremely low in nutrients and water-holding capacity and high in acidity and aluminum—the vegetation consists of many tree species that grow nowhere else. In 1974, Janzen (*1*) proposed that the main factor constraining plants from growing on tropical white-sand soils is attack by herbivorous predators rather than the inability of plants to sustain positive growth under low nutrient conditions. On page 663 of this issue, Fine *et al*. (*2*) report results of the first real field test of this hypothesis. Their findings provide evidence that herbivores influence tropical forest diversity by contributing to habitat specialization (*3*).

One pathway by which herbivores influence plant diversity is by decimating plant populations, particularly at the seed and seedling stage (*4*). This decimation is perhaps most common where many of the herbivores are large, as in tropical Africa and the Indian subcontinent. Mega-herbivores in these regions—including elephants, giraffes, antelope, wildebeests, and rhinos—not only change the number and identity of plant species but convert forest vegetation to grasslands (*5*).

Herbivores need not be large, however, to influence plant diversity. The second pathway by which herbivores influence tropical plant diversity, and plant diversity in general, is by shifting the competitive balance among plant species. If herbivores, including insects, preferentially attack seeds, seedlings, and juveniles of the most competitive species or the most abundant species (frequency-dependent attack), poorer competitors or less frequent plant species could be maintained in the system. For example, the Janzen-Connell hypothe-

The author is in the Department of Biology, University of Missouri–St. Louis, St. Louis, MO 63121, USA. E-mail: robert_marquis@umsl.edu
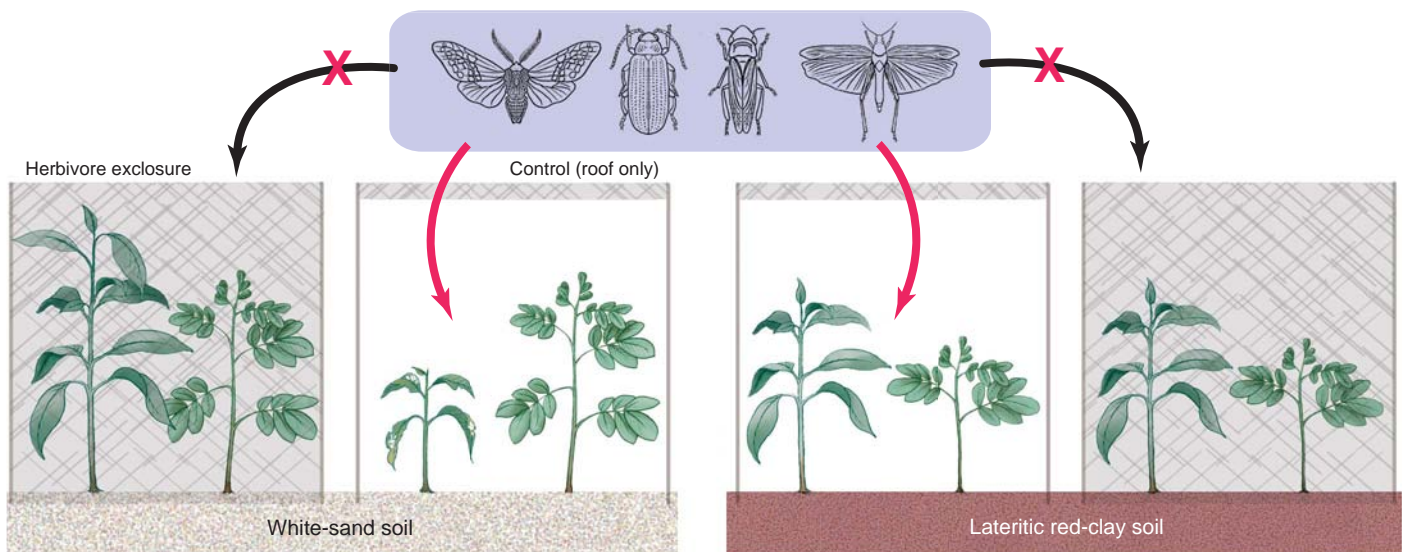
Reciprocal transplant of tropical tree seedlings. Twenty species of tropical tree seedlings (two species are shown) were reciprocally transplanted onto white-sand soils and lateritic red-clay soils. One-half of the tree species used in this experiment naturally grow only on clay soils, whereas the other half are white-sand soil specialists (*2*). Exclosures, which prevented insect herbivores from gaining access to the seedlings, were erected around one-half of the seedlings of each species planted on each soil type. A control exclosure (roof only) was erected around the remainder of the transplanted seedlings, which thus were vulnerable to attack by insect herbivores. Seedling survival and growth were monitored in response to soil source type, soil type at the site of transplanting, and exposure to herbivores.

sis (*6*, *7*) contends that specialist herbivores preferentially attack offspring that are close to parent trees, enabling dispersed offspring to escape. The influence of insect herbivores on the survival of seedlings (*8*) and patterns of density-dependent mortality (*9*) in tropical forests support this hypothesis.

A third mechanism by which herbivores might promote tropical forest diversity is by causing or reinforcing habitat specialization. This is the predicted outcome of Janzen's 1974 hypothesis (*1*). Nutrient availability is so low on these nutrient-poor soils that loss of plant tissue to herbivore attack, and the attendant loss of nutrients, places a poorly defended plant genotype at a great disadvantage. Thus, there would be strong selection for a robust, but costly, anti-herbivore defense in plants growing on nutrient-poor soils. In the presence of herbivores, no species growing in surrounding forest regions on more nutrient-rich soils—clay soils in the Fine *et al*. study (*2*)—would invade white-sand soils because these species would not be sufficiently well-defended to protect themselves against nutrient loss associated with herbivore attack. In the absence of herbivores, however, these species could invade nutrient-poor white-sand soils because, without herbivores, the balance is tipped against white-sand soil specialist species. Likewise, white-sand soil specialists could not invade clay soils because the cost of defense investment has been too high to allow them to compete against species adapted to more nutrient-rich soils.

How might one test this hypothesis? An obvious experiment is to reciprocally transplant both kinds of plant species to both soil types and quantify their success in the presence and absence of herbivores (see the figure). Fine and colleagues (*2*) have completed just such an experiment in the Peruvian Amazon. At the study site, they transplanted 20 species of tree seedlings from six genera, paired such that each pair comprised a clay soil specialist species and a white-sand soil specialist species. They then erected herbivore exclosures consisting of a fine mesh that excluded insect herbivores from one-half of the seedlings in each habitat.

The results of this experiment demonstrate that herbivores help to maintain habitat segregation of the different tree species. On white-sand soils in the absence of herbivores, clay soil specialists survived better and grew more than did white-sand soil specialists. No longer burdened by the loss of resources to herbivores, they outpaced white-sand species whose growth was presumably constrained by their higher investment in antiherbivore defenses. In contrast, when herbivores were free to attack both types of tree species growing on white-sand soils, the white-sand soil specialists fared much better. These results suggest that in the absence of herbivores, clay soil specialists could invade white-sand soils, perhaps even displacing white-sand soil specialists. Meanwhile, on clay soils, white-sand soil species did not grow as well as did clay soil species, again presumably because they were constrained by their

high investment in defense against herbivory. Thus, a large allocation of resources to defense represents a burden to white-sand soil species when on clay soils, but a necessary evil when they are on white-sand soils inhabited by herbivores.

Taking an evolutionary perspective regarding these biotic interactions raises the question of whether herbivores act not only as maintainers but also as promoters of high species diversity. Fine *et al*. (*2*) suggest that herbivores may have promoted diversification through parapatric speciation (where new species form across environmental discontinuities) by accentuating selective gradients among habitats. Here, new species would form on white-sand soils as a result of strong selection by herbivore-soil interactions in the face of continuous gene flow. By determining the evolutionary tree for clay soil and white-sand soil specialists, one could deduce which traits evolved first as white-sand soils were colonized. For example, did defense traits evolve first or did those more directly related to plant-soil relationships, such as special morphologies of plant roots or traits related to interactions with fungi in mycorrhizae?

The Fine *et al*. study provides strong support for the idea that herbivores contribute to the maintenance of high species diversity in local areas that vary in soil type. Herbivores are probably not the sole factor, however. Efforts such as the Fine *et al*. study suggest the need for a holistic approach to solving questions relating to community ecology. A parochial focus on

abiotic or biotic factors alone is likely to provide only limited answers. In addition to soils, the impact of herbivores on tropical forests may vary with elevation and along gradients or discontinuities in soil flooding (*10*), light (*11*), and fire (*12*). But for now, the Fine *et al.* work adds to the mounting evidence that herbivory is a major factor determining the plant composition of tropical forests.

References
1. D. H. Janzen, *Biotropica* **6**, 69 (1974).
2. P. V. A. Fine, I. Mesones, P. D. Coley, *Science* **305**, 663 (2004).
3. R. J. Marquis, in *Biotic Interactions in the Tropics*, D. Burslem, M. Pinard, S. Hartley, Eds. (Cambridge Univ. Press, Cambridge, in press).
4. S. M. Louda, *Ecol. Monogr.* **52**, 25 (1982).
5. A. R. E. Sinclair, P. Arecese, Eds., *Serengeti II, Dynamics, Management, and Conservation of an Ecosystem* (Univ. of Chicago Press, Chicago, IL, 1995).
6. D. H. Janzen, *Am. Nat.* **104**, 501 (1970).
7. J. H. Connell, in *Dynamics of Populations*, P. J. den Boer, G. R. Gradwell, Eds. (Centre for Agricultural Publishing and Documentation, Wageningen, Netherlands, 1971), pp. 298–312.
8. L. A. Hyatt *et al.*, *Oikos* **103**, 590 (2003).
9. K. E. Harms *et al.*, *Nature* **404**, 493 (2000).
10. R. T. King, *Biotropica* **35**, 462 (2003).
11. S. J. DeWalt *et al.*, *Ecology* **85**, 471 (2004).
12. H. T. Dublin, in *Serengeti II, Dynamics, Management, and Conservation of an Ecosystem*, A. R. E. Sinclair, P. Arecese, Eds. (Univ. of Chicago Press, Chicago, 1995), pp. 71–90.

GEOLOGY

A New Period for the Geologic Time Scale

Andrew H. Knoll, Malcolm R. Walter, Guy M. Narbonne, Nicholas Christie-Blick

The geologic time scale stands as a major achievement of 19th-century science, a coherent record of our planet's history fashioned from myriad details of individual rock outcroppings. The eras, periods, and finer divisions of the scale not only codify geologic time, they reflect our accumulated understanding of Earth's past—or at least its more recent past. The Cambrian Period, with its fossil record of animal diversification, began only 543 million years ago (Ma), when Earth was already 4000 million years old (see the figure). In the 19th century and for much of the 20th century, the beginning of the Cambrian (also the beginning of the Paleozoic era and the Phanerozoic eon) marked the most distant temporal reaches of Earth's tractable historical record. The absence of skeletonized fossils that mark Phanerozoic time made Precambrian rocks difficult to correlate, and so the fine stratigraphic divisions of the younger record gave way to broad intervals that permitted only limited insight into foundational events of Earth history. In 1991, perhaps out of resignation, the International Union of Geological Sciences (IUGS) approved a division of Precambrian time into eons, eras, and periods defined strictly by chronometric age, without reference to events recorded in sedimentary rocks (*1*). The eras stuck, but the proposed period names are seldom used.

A. H. Knoll is in the Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA. M. R. Walter is at the Australian Centre for Astrobiology, Department of Earth and Planetary Sciences, Macquarie University, Sydney, NSW 2109, Australia. G. M. Narbonne is in the Department of Geological Sciences and Geological Engineering, Queen's University, Kingston, Ontario K7L 3N6, Canada. N. Christie-Blick is in the Department of Earth and Environmental Sciences and Lamont-Doherty Earth Observatory of Columbia University, Palisades, NY 10964, USA. E-mail: aknoll@oeb.harvard.edu

This tradition was swept aside in March this year with the approval by IUGS of an addition to the geologic time scale: the Ediacaran Period (*2*). This newly ratified period, which directly precedes the Cambrian, is the first Precambrian interval to be defined according to the principles that govern the Phanerozoic time scale. It is also the first stratigraphically defined new period of any sort to be added since 1891 when Williams divided the Carboniferous Period in two (Mississippian and Pennsylvanian). The distinctive character of the Ediacaran interval has been recognized for decades, and numerous geologists—including Sokolov, Termier and Termier, and Cloud and Glaessner (*2*)—have proposed formal definitions of this interval. Now, in accordance with international rules, the new period has been defined by an event recorded in a single section of rock outcropping termed the global stratotype section and point (GSSP). (The GSSP is the reference section that defines the "standard" for recognition of the base of the new period worldwide.) The initial GSSP of the Ediacaran Period lies at the base of a texturally and chemically distinctive carbonate layer that overlies glaciogenic rocks in an exposure along Enorama Creek in the Flinders Ranges, South Australia (*2*) (see the figure). The period's end coincides with the beginning of the Cambrian Period, which is defined by its own initial GSSP residing in Newfoundland, Canada.

Formalisms aside, international ratification of the new period reflects our expanding knowledge of Earth's deep physical and biological history. The Ediacaran Period, in fact, constitutes a distinct chapter in that history, bounded below by global ice ages and above by the diversification of animal life—and characterized most vividly by the unusual, mostly soft-bodied fossils that give it its name. The unique morphologies of the Ediacara biota have spawned widely varying systematic interpretations—from giant protists and lichens to seaweeds and extinct experiments in multicellularity. Most paleontologists, however, agree that the assemblage includes early cnidarian-grade animals, as well as burrows and trails and perhaps body fossils of early bilateral organisms (bilaterians) (*3*).
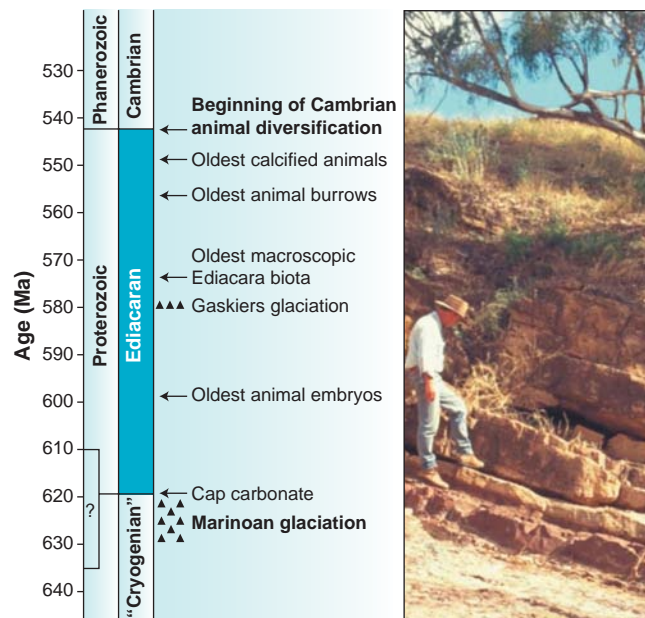
Looking for a few good rocks. (Left) Major events associated with the Ediacaran Period. Brackets indicate uncertainty in the chronometric age of the GSSP. **(Right)** The formally defined base (GSSP) of the Ediacaran Period in Enorama Creek, Australia, is located at the contact of Marinoan glacial rocks and overlying Ediacaran cap carbonates (at the right foot of the geologist).

Dates are important. The beginning of the period remains to be determined precisely, but the uranium-lead (U-Pb) zircon dating method gives a maximum age of 635.5 ± 1.2 Ma for zircons from volcanic ash within glacial diamictites in Namibia (*4*). Meanwhile, a Pb-Pb date of 599 ± 4 Ma for postglacial phosphorites from China (*5*) provides a minimum age for the beginning of the Ediacaran Period. The earliest known animal fossils—microscopic eggs, embryos, and segmented skeletal tubes—are found in the phosphorites (*6*). Following one last, regionally distributed glaciation, moderately diverse macroscopic fossils appear in ~575 Ma rocks from Newfoundland (*7*). Bilaterian animal trails enter the record no later than 555 Ma, and calcified skeletons (of a distinctively Ediacaran, not Cambrian, aspect) by 549 Ma (*8*). Ediacaran assemblages persisted until the end of the period, separated from Cambrian diversification by a major, short-lived perturbation in the carbon isotopic record.

If Ediacaran fossils characterize the period, why don't they define it? The simple answer is that the fossils are scarce and, consequently, there are large uncertainties regarding correlation. Among sedimentary basins, the first appearance of Ediacara-type fossils can differ by 10 million years or more. This is why the Ediacaran Period departs rather abruptly from Phanerozoic convention in defining the beginning of the period by a climatic/geochemical event. The unusual depletion of ^{13}C in the texturally striking carbonates that veneer Marinoan glacial rocks is recognized globally and widely accepted as a paleoceanographic signature of rapid deglaciation, although mechanistic interpretations differ (*9*, *10*). More generally, large secular variations in the isotopic compositions of carbon, sulfur, and strontium have come to play an important part in the correlation of Neoproterozoic (1000 to 543 Ma) sedimentary rocks. This works well because younger Proterozoic strata record huge sec-

ular variations in the composition of seawater that reflect not only global ice ages, but also biospheric oxidation and global tectonic events. Indeed, the Neoproterozoic has emerged as a primary focus of Earth systems history, as scientists seek to understand the complex interactions between planet and life that gave rise to the Phanerozoic world. Testifying to this effort, the new Ediacaran Period provides a first extension of the geologic time scale into Earth's Precambrian past. It will not be the last.

References

1. K. A. Plumb, *Episodes* **14**, 139 (1991).
2. A. H. Knoll *et al.*, *Lethaia*, in press.
3. G. M. Narbonne, *GSA Today* **8**, 1 (1998).
4. K.-H. Hoffmann *et al.*, *Geology*, in press.
5. G. H. Barfod *et al.*, *Earth Planet. Sci. Lett.* **201**, 203 (2002).
6. S. Xiao, A. H. Knoll, *J. Paleontol.* **74**, 767 (2000).
7. G. M. Narbonne, J. G. Gehling, *Geology* **31**, 27 (2003).
8. M. W. Martin *et al.*, *Science* **288**, 841 (1998).
9. P. F. Hoffman, D. P. Schrag, *Terra Nova* **14**, 129 (2002).
10. G. Q. Jiang *et al.*, *Nature* **426**, 822 (2003).

PLANETARY SCIENCE

A Unique Chunk of the Moon

Randy L. Korotev

In 1982 a team of U.S. scientists collecting meteorites in Antarctica found a fragment of the Moon. The 31-g meteorite, now called Allan Hills (ALHA) 81005, had once been a rock or a piece of a rock that existed at or near the Moon's surface. At some time in the past, a meteoroid collided with the Moon and accelerated the rock to lunar escape velocity. After orbiting Earth for less than 200,000 years, the rock was captured by Earth's gravitational field, landed in Antarctica, and was buried by snow. There it became a miniscule part of a huge glacier, which also carried other meteorites that had fallen over the years. The glacier's flow is impeded by the Transantarctic Mountains, and near the mountains meteorites are continually exposed at the surface as wind and sun ablate and sublime away the ice that encases them. The collecting team immediately recognized that ALHA 81005 did not look like the other meteorites that they were collecting, all of which were fragments of asteroids. Meteorite curators at the NASA Johnson Space Center, having seen a lot of Moon rocks from the Apollo missions, suspected that it was a Moon rock. Further studies have confirmed their suspicion (*1*). The stone was the first to be recognized as a lunar meteorite,

although three others not yet classified had been collected in Antarctica 3 years earlier by a team from the Japanese National Institute of Polar Research. Since 1979, about 30 lunar meteorites have been found, all in deserts. On page 657 of this issue, Gnos *et al.* (*2*) describe the most unique lunar meteorite found to date. This 206-g stone, known as Sayh al Uhaymir (SaU) 169, was found in the Sultanate of Oman in January 2002.

On the basis of the wide ranges in composition, mineralogy, texture, and cosmic-ray exposure ages, the 30 lunar meteorites likely represent at least 20 impacts on the lunar surface, although the crater of origin is not known for any of them. For any given lunar meteorite, the fact that we don't know where on the Moon it originates is a serious detriment to geologic interpretation of data derived from the stone. However, the meteorites are samples from many random locations, and this characteristic provides important information not available from the Apollo samples, all of which were collected on six missions to the central nearside (see the figure).
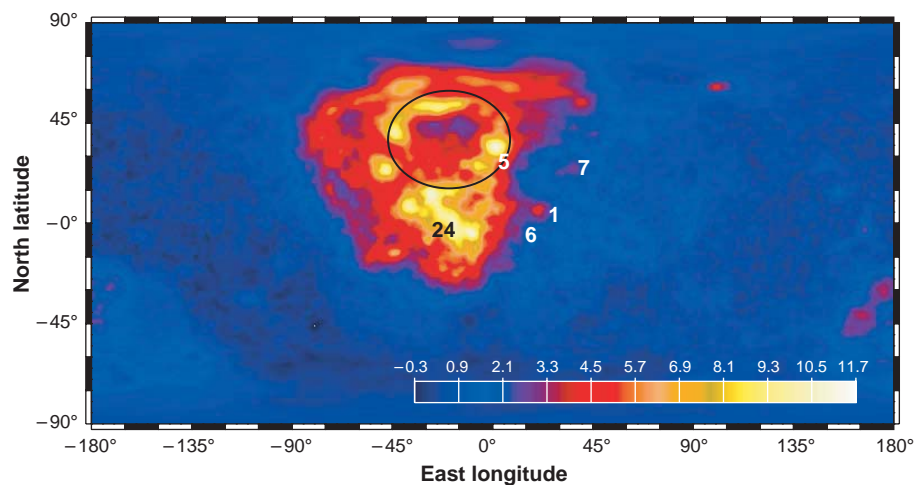
Rich in radioactive elements. Distribution of thorium on the lunar surface [adapted from (*5*)]. Color scale shows thorium abundance in µg/g. The numbers represent the locations of the six Apollo landing sites (1 = Apollo 11, 2 = Apollo 12, and so on; landing sites 2 and 4 are adjacent). The ellipse indicates the position of the Imbrium basin. The center of the figure is the center of the nearside, as viewed from Earth. Most of the Moon's thorium and other incompatible elements are concentrated in the northwest quadrant of the nearside.

The author is in the Department of Earth and Planetary Sciences, Washington University in St. Louis, St. Louis, MO 63130, USA. E-mail: korotev@wustl.edu

It came as a surprise to early lunar sample researchers that samples from the Apollo 12 and 14 missions had very high concentrations of the suite of trace chemical elements that geochemists categorize as "incompatible." On the Moon, incompatible elements include phosphorus, the rare earth elements, and the three most important naturally occurring radioactive elements, potassium, thorium, and uranium. Lunar rocks consist mainly of four minerals. When those minerals crystallize from a magma, the incompatible elements are excluded from the solid phases and are concentrated in the liquid phase. The existence of lunar rocks with high concentrations of incompatible elements indicated that igneous differentiation of the Moon had occurred to an advanced degree and that the Moon, unlike the parent bodies of most meteorites, was not a primitive object.

Partial mapping of the lunar surface by gamma-ray spectrometers aboard Apollo orbiting command modules showed that the region in the vicinity of the Apollo 12 and 14 sites was rich in radioactive elements. However, it was not until 1998 that the first global geochemical coverage of the Moon was obtained by gamma-ray and neutron spectrometers aboard the Lunar Prospector mission (3–5). It only became apparent more than 25 years after the last Apollo mission that three of the missions, Apollos 12, 14, and 15, had landed in a region that was uniquely rich in incompatible elements compared to most regions of the Moon and that the other three missions were in near proximity to that geochemically anomalous region (see the figure).

Various lines of evidence suggest that the early Moon was largely molten. The Lunar Prospector data showed that the chemical differentiation of the Moon, as it formed its core, mantle, and crust, was asymmetric. The last liquid, by then rich in incompatible trace elements, concentrated in what is now the northwest quadrant of the nearside. One of the last and largest basin-forming impactors, the one that produced the Imbrium basin, struck this geochemically anomalous region 3.9 billion years ago, spreading thorium-rich ejecta over the surface of the Moon (6). All six Apollo landing sites contain rocks, mainly ancient impact-melt breccias, that are rich in thorium (typically 8 to 20 µg/g) and other incompatible elements. One interpretation is that these thorium-rich impact melt breccias were produced when the Imbrium impactor struck the incompatible-element–rich region (7). Another is that a cataclysmic set of impact events occurring within a short interval about 3.9 billion years ago produced all of the major nearside lunar basins and that several of those

impacts excavated high-thorium material (8). Unlike typical rocks from the Apollo sites, most lunar meteorites, including AL-HA 81005, have low concentrations of thorium, typically less than 1 µg/g. These low-thorium meteorites must originate from the vast portions of the Moon, mainly on the farside, with low surface concentrations of thorium (9).

SaU 169 represents the opposite extreme. It is an impact-melt breccia with exceedingly high concentrations of thorium (33 µg/g) and other incompatible elements. As Gnos et al. argue, SaU 169 almost certainly originates from within the high-thorium anomaly. If lunar meteorites are random samples of the Moon, it was predictable that sooner or later a high-thorium lunar meteorite would be found. It is nevertheless ironic that even though the Apollo missions inadvertently visited some of the most thorium-rich areas of the Moon, SaU 169 is richer in incompatible elements than any rock-sized Apollo sample.

The most significant aspect of the work of Gnos et al. is the ^{207}Pb-^{206}Pb crystallization age of 3.909 ± 0.009 billion years that they obtain for the impact melt on the basis of ion microprobe analysis of zircons. They

conclude that this age precisely dates the Imbrium impact, although the age is significantly older than the previous best working ages of 3.85 ± 0.02 billion years (8) for Imbrium and 3.89 ± 0.01 billion years for the Serenitatis basin (10). The data for SaU 169 call into question just which isotopic systems best record the crystallization age of an impact melt and whether the small differences (<2%) in ^{40}Ar-^{39}Ar ages among ancient impact-melt rocks are significant. It is now imperative that ^{207}Pb-^{206}Pb ages be obtained from zircons in thorium-rich melt breccias from the Apollo landing sites for comparison.

References

1. U. B. Marvin, *Geophys. Res. Lett.* **10**, 775 (1983).
2. E. Gnos *et al.*, *Science* **305**, 657 (2004).
3. D. J. Lawrence *et al.*, *Science* **281**, 1484 (1998).
4. R. C. Elphic *et al.*, *J. Geophys. Res.* **105**, 20333 (2000).
5. D. J. Lawrence *et al.*, *J. Geophys. Res.* **108**, 5102, 10.1029/2003JE002050 (2003).
6. L. A. Haskin, *J. Geophys. Res.* **103**, 1679 (1998).
7. L. A. Haskin *et al.*, *Meteorit. Planet. Sci.* **33**, 959 (1998).
8. G. Ryder, *J. Geophys. Res.* **107**, 6-1, 10.1029/2001JE001583 (2002).
9. R. L. Korotev *et al.*, *Geochim. Cosmochim. Acta* **67**, 4895 (2003).
10. G. B. Dalrymple, G. Ryder, *J. Geophys. Res.* **101**, 26069 (1996).

MATERIALS SCIENCE

Watching the Nanograins Roll

E. Ma

Conventional engineering materials are usually polycrystalline solids composed of crystallites (grains) that are many micrometers in diameter (d). The recent advent of nanocrystalline materials, for which d is less than 100 nm, has opened new opportunities for research and applications. For the "upper nano" regime with $d \approx$ 100 nm, research on mechanical properties has focused on tailoring the grain and/or boundary structures for optimized strength and ductility (1). The "lower nano" regime with $d < \sim$30 nm, on the other hand, is the realm for discovery of new deformation mechanisms (2–13). This is because the nucleation and movement of line defects called dislocations—the main carrier of plastic deformation in coarse-grained metals—is projected to be difficult in such tiny grains where the distance between dislocation pinning points becomes very small, demanding very high stresses to activate dislocation sources. Deformation processes governed by abundant grain boundaries

The author is in the Department of Materials Science and Engineering, Johns Hopkins University, Baltimore, MD 21218, USA. E-mail: ema@jhu.edu

may become important, as demonstrated in nanocrystalline nickel by Shan *et al.* on page 654 of this issue (2). Using an in situ dark-field transmission electron microscopy (TEM) technique, the authors recorded the frequent rotation of nanocrystals ($d \approx$ 6 nm) into larger aggregates of neighboring grains during deformation.

In general, a deforming grain is forced to rotate in response to the external stresses exerted upon it by its neighbors. For conventional metals, grain rotation during deformation often accompanies the formation of texture (preferred orientation) and is accomplished by the microscopic dislocation glide on multiple active slip systems in the grains. Such extensive dislocation activity is unlikely in 6-nm grains, and texturing was not observed in heavily cold-rolled nanocrystalline palladium (12). Grain rotation can also be caused by extensive grain boundary sliding and diffusion, which usually occur only at elevated temperatures. But molecular dynamics (MD) computer simulations predict that such a mechanism will become operative and even dominant when d is reduced to ~10 nm (4, 7, 8). Evidence for this has been sought for many

years, but direct experimental confirmation remained elusive except in extremely thin (~20 nm) films used for high-resolution TEM experiments (*3*).

The experimental findings of Shan *et al.* (*2*) are also valuable because our understanding of the deformation mechanisms of nanocrystalline metals has relied heavily on MD simulations. With high loads and extreme strain rates applied to produce a measurable strain within the subnanosecond MD time scale, the simulations may exclude certain time-dependent processes
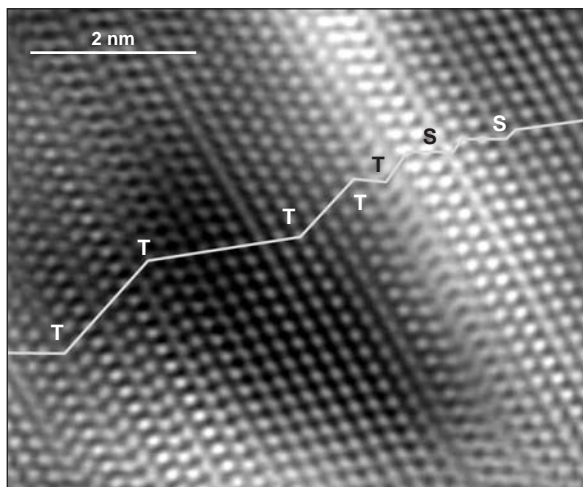
Deformation debris. An example of features left by partial dislocation–mediated processes during deformation [adapted from (*6*)]. A Fourier-filtered high-resolution TEM shows the stacking faults (S) and deformation twins (T) in a deformed nanocrystalline aluminum grain (*6*). Note the mirror symmetry across the twin boundaries, where deposited partial dislocations are also observed.

and cannot determine the true rate-limiting processes in real-world experiments. The MD results are sources of inspiration and guidance but do not directly validate or disprove the existence of a mechanism (*4*, *9*).

Taken together, recent TEM observations (*2*, *4–6*), in situ synchrotron x-ray diffraction (XRD) experiments (*10*), and advances in MD simulations begin to paint a unifying picture for the unusual deformation mechanisms in nanocrystalline face-centered-cubic metals (such as nickel and copper) (*7*, *11*). When dislocations are still involved, their operation takes different forms. First, the intragrain Frank-Read dislocation sources dominant in conventional polycrystals cease to control deformation. Instead, dislocations are nucleated out of grain boundary (GB) sources. There is also indirect evidence, including the increased strain rate sensitivity and unusually low activation volume, that hints at the defect (boundary)–assisted nucleation of dislocations as the thermally activated rate-controlling process (*14*). For *d* on the order of 30 nm, the dislocations traverse the grain and

disappear into the opposing GBs, with little chance of storage inside the grains (*4*, *5*). Second, partial dislocation emission from a GB becomes more favored at sufficiently small *d*. Indeed, estimates showed that the energy cost for the nucleation of one partial dislocation at a time, which is observed in MD simulations for all the nanocrystalline metals studied (including those with high stacking fault energy) (*7*, *9*), can be lower than that for emitting a full (perfect) dislocation (*6*). The partial dislocation–mediated processes do leave behind debris, such as the stacking faults and deformation twins in the figure (*6*, *12*, *15*), that can be observed in postmortem TEM. However, the nucleation of a trailing partial dislocation on the same plane of an existing leading partial dislocation, before the latter gets absorbed by the opposing GB, is often possible and can be an energetically less expensive process than other options [such as the emission of a second partial dislocation on the adjacent plane for twin nucleation (*9*)]. The trailing partial dislocation then erases the stacking fault and may catch up with the leading partial dislocation. The resulting full dislocations are not often observed in TEM, as they leave no footprint in their wake after traversing the grain. Their action was inferred from the in situ XRD measurements (*10*). It is interesting that a unit dislocation trapped inside the grain under loading has been captured by Shan *et al.* (*2*).

Large stresses are required to drive deformation in nanoscale grains. The local stress intensity can be particularly high and varied depending on loading conditions, such that one may observe all types of slip: extended partial dislocations forming stacking faults, full dislocations, and deformation twinning (*6*, *9*). Note that dislocation activities persist even when *d* is reduced down to ~10 nm (*2*, *6*, *12*, *15*). At such small grain sizes, GB sliding and grain rotation become detectable, concurrent with dislocation activities and possibly also accommodated by atomic shuffling at the GBs (*4*). A supporting finding for the active role of grain rotation and GB sliding, as directly observed by Shan *et al.*, is the lack of crystallographic texture in nanocrystalline palladium grains that remained equiaxed even after large plastic deformation (*12*). Rosner *et al.* also suggest that the number of glide systems active in the coplanar twin-

ning they observed is obviously short of the five slip systems required for a general deformation in polycrystals (*12*).

Simulations with idealized samples have predicted predominant GB sliding at *d* < ~10 nm (*7*), but the experimental samples available so far [see figure 1 of Shan *et al.* (*2*)] always contain a grain size distribution and some impurities, with a large (number and volume) fraction of the grains larger than 10 nm. Some deposited grains in columnar shape also may not be conducive to GB sliding and grain rotation. As a result, GB–mediated plasticity may be a contributing but not yet dominant mechanism. The Hall-Petch relationship (increasing strength with decreasing *d*) may continue to hold for the nickel films of Shan *et al.* (*13*) without displaying an obvious maximum strength beyond which an inverse Hall-Petch behavior (softening) takes over (*7*).

Grain rotation and GB sliding, with no evidence of dislocation activities, were thought to control deformation in gold films with *d* = 10 nm (*3*), but the sample used was ultrathin (10 to 20 nm) and the deformation rate was very slow. In this case, the two-dimensional geometry lacks the three-dimensional constraints and has a very high surface-to-volume ratio. This accentuates the thermally activated processes facilitating grain rotation, such as diffusion, especially when under electron irradiation in TEM. Note that even the thicker films of Shan *et al.* are not free of such effects (*16*), even though the authors believe that the surface effects and stress-assisted grain growth (due to driving forces to reduce surface or GB energy) are negligible (*2*). Therefore, although the result of Shan *et al.* brings new insight into the deformation of extremely small grains, caution should be exercised before generalizing the behavior in TEM foils as fully representative of bulk deformation.

References and Notes
1. Y. M. Wang *et al.*, *Nature* **419**, 912 (2002).
2. Z. Shan *et al.*, *Science* **305**, 654 (2004).
3. M. Ke *et al.*, *Nanostruct. Mater.* **5**, 689 (1995).
4. K. S. Kumar, H. Van Swygenhoven, S. Suresh, *Acta Mater.* **51**, 387 (2003).
5. R. C. Hugo *et al.*, *Acta Mater.* **51**, 1937 (2003).
6. M. W. Chen *et al.*, *Science* **300**, 1275 (2003).
7. J. Schiotz, K. W. Jacobsen, *Science* **301**, 1357 (2003).
8. V. Yamakov *et al.*, *Nature Mater.* **3**, 43 (2004).
9. H. Van Swygenhoven *et al.*, *Nature Mater.* **3**, 399 (2004).
10. Z. Budrovic *et al.*, *Science* **304**, 273 (2004).
11. S. Cheng *et al.*, *Acta Mater.* **51**, 4505 (2003).
12. H. Rosner *et al.*, *Philos. Mag. Lett.* **84**, 321 (2004).
13. J. A. Knapp, D. M. Follstaedt, *J. Mater. Res.* **19**, 218 (2004).
14. Our stress relaxation and jump tests for nickel with *d* = 30 nm showed a room-temperature strain rate sensitivity four times that of coarse-grained nickel and an activation volume as small as 7 to 20 b^3, where *b* is the Burgers vector.
15. X. Z. Liao *et al.*, *Appl. Phys. Lett.* **84**, 592 (2004).
16. P. M. Derlet *et al.*, *Philos. Mag. A* **82**, 1 (2002).

REVIEW

The Pathophysiology of Mitochondrial Cell Death

Douglas R. Green[1]* and Guido Kroemer[2]*

In the mitochondrial pathway of apoptosis, caspase activation is closely linked to mitochondrial outer membrane permeabilization (MOMP). Numerous pro-apoptotic signal-transducing molecules and pathological stimuli converge on mitochondria to induce MOMP. The local regulation and execution of MOMP involve proteins from the Bcl-2 family, mitochondrial lipids, proteins that regulate bioenergetic metabolite flux, and putative components of the permeability transition pore. MOMP is lethal because it results in the release of caspase-activating molecules and caspase-independent death effectors, metabolic failure in the mitochondria, or both. Drugs designed to suppress excessive MOMP may avoid pathological cell death, and the therapeutic induction of MOMP may restore apoptosis in cancer cells in which it is disabled. The general rules governing the pathophysiology of MOMP and controversial issues regarding its regulation are discussed.

The major form of apoptosis seen in most settings in vertebrate cells proceeds through the mitochondrial pathway, defined by a pivotal event in the process—mitochondrial outer membrane permeabilization (MOMP). MOMP occurs suddenly during apoptosis (*1*), leading to the release of proteins normally found in the space between the inner and outer mitochondrial membranes (including cytochrome c, AIF, and others). Before, during, or after MOMP, there is frequently a dissipation of the mitochondrial inner transmembrane potential ($\Delta\Psi$m), and the timing of MOMP versus $\Delta\Psi$m loss can provide clues to the mechanism involved in a particular setting. MOMP precipitates the death of the cell through as many as three general mechanisms, including the release of molecules involved in the activation of caspases that orchestrate downstream events often associated with apoptosis, the release of molecules involved in caspase-independent cell death, and the loss of mitochondrial functions essential for cell survival (table S1). A pathophysiological role for MOMP is emerging.

Mechanisms of MOMP and the Decision to Die

The mechanisms responsible for MOMP during apoptosis remain controversial, although it is clear that many proteins can inhibit or prevent MOMP by local effects on mitochondrial membranes (tables S2 and S3). In general, two classes of mechanism have been described and each may function under different circumstances: those in which the inner mitochondrial membrane participates, and those involving only the outer membrane (Fig. 1).

In the first class of mechanism, a pore opens in the inner membrane, allowing water and molecules up to ~1.5 kD to pass through. Although most models of this pore, the permeability transition (PT) pore, postulate roles for the adenine nucleotide transporter (ANT) in the inner membrane and the voltage-dependent anion channel (VDAC) in the outer membrane (*2*), this is a hypothetical model (supporting online text). Recent evidence has shown that the PT pore can form in the absence of the ANT (*3*), and alternative models accounting for this pore have been proposed (*4*). Opening of the PT pore can be triggered by multiple stimuli and leads to (i) $\Delta\Psi$m loss as ions equilibrate across this membrane, and (ii) swelling of the matrix as water enters. The latter can result in sufficient swelling to break the outer membrane to produce MOMP. It should be noted that although loss of $\Delta\Psi$m accompanies irreversible PT, many other events can produce this loss. Loss of $\Delta\Psi$m is not sufficient to prove the involvement of PT. Conversely, PT pore opening can be transient (through flickering of the pore), and therefore sustained $\Delta\Psi$m does not provide a firm argument against the involvement of the PT pore unless monitored continuously throughout MOMP. In view of the difficulties of quantifying PT in living cells, and in the absence of a clear molecular substrate for the pore (supporting online text), it may be a pragmatic approach to define PT-associated MOMP as a process that can be inhibited by some ligands of putative PT pore constituents

such as VDAC, ANT, or the ANT-interacting protein cyclophilin D (a target of cyclosporin A) (*5*). Ideally, methods that directly assess the permeability of the inner membrane (*6*) should be employed.

The second class of mechanism for MOMP does not involve a major role for PT or the mitochondrial inner membrane. This appears to be mediated by members of the Bcl-2 family of apoptosis-regulating proteins acting directly on the outer mitochondrial membrane (table S4). Anti-apoptotic Bcl-2 family members function to block MOMP, whereas the various pro-apoptotic members promote it. The latter fall into two general subfamilies, based on sharing of Bcl-2 homology (BH) domains. BH123 (or multidomain) proteins share BH1, BH2, and BH3 and appear to be effectors of MOMP, because cells from mice lacking the two major BH123 proteins, Bax and Bak, fail to undergo MOMP in response to a wide range of apoptotic signals (*7*). The other subfamily, the BH3-only proteins (that contain only the BH3 domains), can act either to activate Bax and Bak or to interfere with the anti-apoptotic Bcl-2 family members (*8*).

Vesicles composed of purified mitochondrial outer membranes can be permeabilized in response to activated forms of recombinant Bax or Bid, the latter presumably acting through Bak on the outer membrane (*9*). Further, vesicles composed of mitochondrial lipids (without other mitochondrial proteins) were permeabilized by recombinant, monomeric Bax, provided that active recombinant Bid or a BH3 peptide derived from Bid was present. This generated openings of indeterminate size that could not be visualized by conventional ultrastructural techniques. Such openings may be consistent with large lipidic pores composed of activated BH123 proteins and lipids with potential for negative curvature in membranes (*10*), for instance, the mitochondrial lipid cardiolipin. However, the presence of cardiolipin in the mitochondrial outer membrane remains controversial.

Some studies have implicated the outer membrane protein VDAC in MOMP. Bax and Bak can bind to VDAC, but possibly with different effects. Although the Bax-VDAC interaction is suggested to cause MOMP, interaction of Bak with VDAC-2 appears to be inhibitory (*11*). One possibility is that VDAC functions to sequester small amounts of cardiolipin or related lipids

[1]Division of Cellular Immunology, La Jolla Institute for Allergy and Immunology, 10355 Science Center Drive, San Diego, CA 92121, USA. [2]Centre National de la Recherche Scientifique, Unité Mixte de Recherche 8125, Institut Gustave Roussy, 39 rue Camille-Desmoulins, F-94805 Villejuif, France.

*To whom correspondence should be addressed. E-mail: doug@liai.org (D.R.G.) and kroemer@igr.fr (G.K.)

present in the outer membrane to microdomains in which local concentrations of these lipids may be sufficient to allow permeabilization of the membrane by activated Bax or Bak. This would also account for the apparent binding of Bax to VDAC.

PT-independent MOMP can be followed by secondary PT. In sympathetic neurons deprived of nerve growth factor, Bax-dependent, PT-independent MOMP associated with cytochrome c release causes caspase activation and apoptosis. However, in the presence of caspase inhibitors, such cells survive until $\Delta\Psi$m drops (12). Studies in which cyclosporin A blocks the $\Delta\Psi$m loss and commitment to death suggest that PT determines the point of no return in these cells (13).

Upon MOMP, proteins of the intermembrane space are released, although whether or not all proteins are released simultaneously remains controversial. One suggestion is that these proteins are differentially sequestered in the intermembrane space and that secondary events are required for the release of some of them (14, 15). For example, remodeling of the matrix and inner mitochondrial membrane may be required for the release of cytochrome c in some cases (15), although in other cases this was not observed (16). Such remodeling has been suggested to be mediated by PT (15). Further, mitochondrial fission can occur around the time of MOMP (17) and proteins that regulate fusion or fission of mitochondria appear to affect which proteins can be released upon MOMP.

Irrespective of its mechanisms, MOMP can seal the point of no return of the lethal process by the release of caspase activators such as cytochrome c (table S1). Once activated, caspases can cause a rapid loss of mitochondrial functions. Upon MOMP, executioner caspases can cleave the NDUSF1 subunit of respiratory complex I, and mutation of its single cleavage site can preserve mitochondrial functions during apoptosis (18). This can delay plasma membrane events associated with caspase activation, including loss of plasma membrane integrity and externalization of phosphatidylserine, thus indicating an important role for disruption of mitochondrial function in apoptotic cell death. Nonetheless, even without caspase activation, MOMP generally results in cell death through the release of multiple caspase-independent death effectors, as well as loss of essential mitochondrial functions (table S1 and supporting online text).

Upstream of MOMP

Multiple distinct signaling pathways converge on MOMP (tables S2 and S3). Although some of the BH3-only proteins in the Bcl-2 family have the capacity to activate Bax and Bak or, conversely, inhibit the anti-apoptotic Bcl-2 family members, other

molecules may have these properties as well. The tumor suppressor p53 acts, in part, to induce apoptosis by inducing expression of the BH3-only protein PUMA, and PUMA-deficient cells display a resistance to p53-mediated apoptosis (19, 20). However, p53 can trigger MOMP and apoptosis in the absence of transcription, and this can occur through direct activation of Bax (21) or Bak (22) or through binding to Bcl-2 and Bcl-X$_L$, which blocks their activity (21, 23). Resolving the role for this mechanism versus that of transcriptional regulation will be important in understanding the apoptotic function of p53.

An emerging theme is one of nuclear proteins functioning in the cytosol through direct interactions with Bcl-2 family proteins. Ku70, involved in DNA repair, can inhibit Bax (24). Another nuclear protein, TR3, binds Bcl-2 and perhaps promotes MOMP through this interaction (25). Histone 1.2, released from the nucleus upon X-ray–induced DNA damage, can trigger MOMP (26), perhaps through an interaction with Bcl-2 family members. Hexokinase can interact with VDAC, and this interaction may inhibit the ability of Bax to cause MOMP (27). Intriguingly, enforced expression of hexokinase together with the glucose transporter Glut-1 is sufficient to confer cell survival (28).

Alternatively, Bcl-2 family members may act independently of mitochondria and upstream of MOMP. Cells lacking Bax and Bak display reduced calcium efflux from the en-
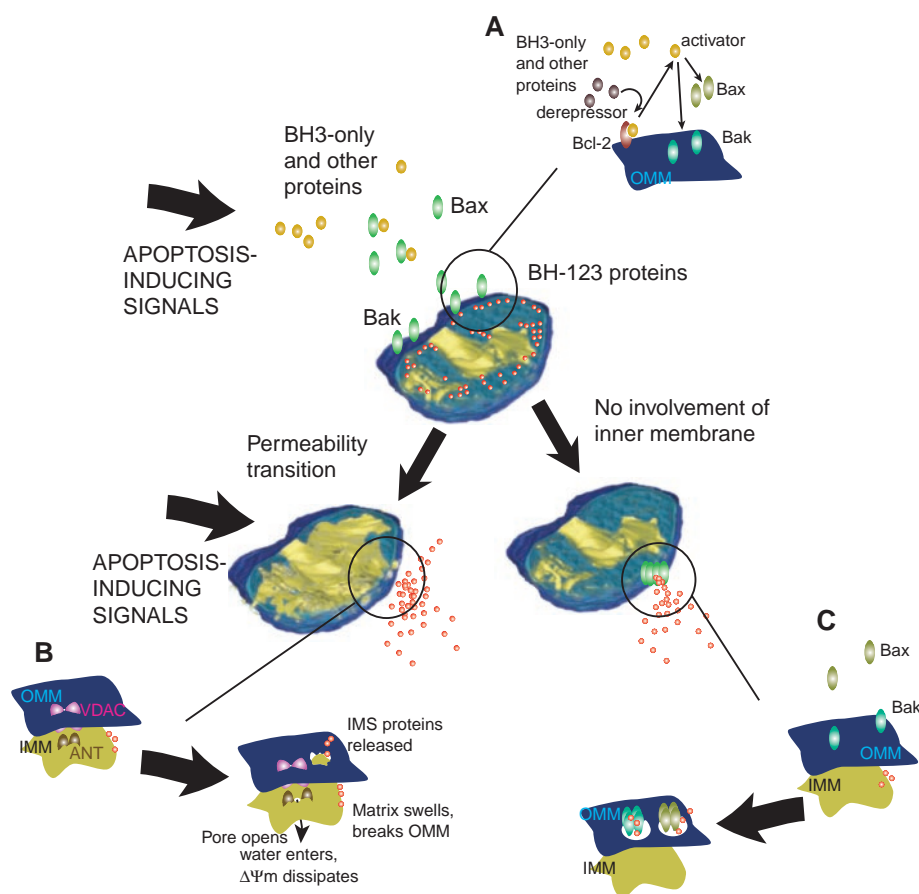
Fig. 1. Mechanisms for MOMP during apoptosis. (**A**) Signals for the induction of apoptosis (top) engage the activities of a subgroup of pro-apoptotic, BH3-only members of the Bcl-2 protein family and other proteins, which in turn activate the pro-apoptotic, BH123 proteins Bax and Bak to oligomerize and insert into the outer mitochondrial membrane (OMM). Other BH3-only proteins can act indirectly by releasing the first subgroup of BH123-activators from the anti-apoptotic Bcl-2 family proteins that sequester them. The BH123 proteins engage either of the two mechanisms that follow, perhaps depending on cell type or other conditions. (**B**) In PT-dependent MOMP, apoptosis-inducing signals act directly or indirectly to open the putative PT pore. This is composed of ANT or other proteins in the inner mitochondrial membrane (IMM) and is associated with VDAC and perhaps other proteins in the OMM. Opening the pore allows water to enter the matrix and ions to equilibrate, dissipating $\Delta\Psi$m at least transiently. The matrix swells, rupturing the OMM to release proteins of the mitochondrial intermembrane space (IMS). (**C**) In PT-independent MOMP, BH123 proteins, perhaps with other proteins, cause the formation of pores in the OMM through which IMS proteins are released. The mitochondrial tomograph was a gift from G. Perkins and D. Newmeyer. This was modified in the illustration.

Table 1. Examples of pathogenic processes involving excessive or deficient MOMP.

Disease	Pathogenic perturbation of MOMP	Pharmacological correction of deregulated MOMP
Ischemia reperfusion damage of brain or heart	Redox stress, excessive Ca^{2+} load, absent adenosine triphosphate and nicotine adenine denucleotide, and accumulating fatty acids favor PT and MOMP.	Bcl-2 inhibitors of the PT pore, as well as mito K^{ATP} channel openers, can exert neuro- or cardioprotective effects.
Neurodegenerative diseases	Respiratory dysfunction affects highly sensitive neurons in the central nervous system, leading to their premature death.	Putative inhibitors of the PT pore (minocyclin, rasagiline, and tauroursodeoxycholic acid) can prevent neurodegeneration.
Liver disease	Hepatotoxins (including bile acid and ethanol) and hepatitis B or C–encoded proteins induce MOMP.	Ursodeoxycholic acid prevents bile acid–induced PT and thus exerts hepatoprotective effects.
Cancer	MOMP-inhibitory proteins from the Bcl-2 family or unrelated proteins (such as Muc1) enhance apoptosis resistance.	Cytotoxic agents targeting Bcl-2–like proteins, PT pore components, and/or mitochondrial lipids enforce MOMP and kill cancer cells.
HIV-1 infection	Vpr, an accessory HIV-1–encoded protein, can act on ANT and Bax to trigger MOMP. This effect is frequently lost because of a mutation in long-term nonprogressors.	HIV-1 protease inhibitors can inhibit MOMP induced by Vpr in isolated mitochondria.

doplasmic reticulum in response to some stimuli (*29*) [whether this is directly an effect of Bax and Bak or an indirect effect of Bcl-2 on the receptor is controversial (*30*)]. The resulting calcium flux may act on mitochondria to produce MOMP independently of Bax and Bak, by induction of PT (*29*).

The Mitochondrial Pathway of Apoptosis in Pathogenesis

MOMP-dependent apoptosis is involved in major pathologies, with far-reaching medical and pharmaceutical implications (Table 1 and table S5). However, the role of MOMP in disease is often inferred from correlative studies, such as when a disease-associated molecule or pharmaceutical agent is shown to have effects on mitochondria, and should therefore be treated with caution. Nevertheless, the manifestations of MOMP (the mitochondrial release of intermembrane proteins, as well as the dissipation of $\Delta\Psi$m) are frequently observed in disease states with increased cell death (*31*).

Many viruses have acquired the capacity to intercept or to activate the principal signal-transducing pathways leading to cell death. Several proteins from pathogenic viruses are targeted to mitochondria and induce MOMP (Table 1 and table S5), and at least one is a virulence factor; a mutation in the human immunodeficiency virus (HIV) Vpr protein that attenuates its MOMP-inducing activity is statistically associated with a reduced risk to develop acquired immunodeficiency syndrome (AIDS) (*31*). Several oncogenic viruses encode MOMP-inhibitory proteins, and in humans these may be involved in the formation of virally induced lymphomas or Kaposi's sarcoma.

In acute pathologies, for instance after ischemia, a combination of increased intracellular Ca^{2+}, reactive oxygen species, and metabolic perturbations can trigger MOMP, which ultimately accounts for cell loss in cardiac infarction and cerebral stroke. This cell loss involves acute necrosis in the ischemic core and a slower apoptosis in the penumbra. The hypocampal CA1 region, extremely vulnerable to ischemia, contains mitochondria with the highest susceptibility to Ca^{2+}-induced MOMP in vitro (*32*). Key modulators of apoptosis, such as p53 and Bax, facilitate ischemia-induced MOMP and neuronal death (*33*), whereas the MOMP inhibitor Bcl-2 can prevent ischemia-induced neuronal apoptosis.

MOMP is also likely to be involved in chronic neurodegenerative diseases. For example, in Huntington's disease, polyglutamine expansions in huntingtin trigger neuronal death, and this aberration correlates with huntingtin-induced mitochondrial abnormalities. Transgenic overexpression of Bcl-2 can prolong the life span of mice carrying an SOD-1 mutation found in patients with familial amyotrophic lateral sclerosis, a degenerative disease affecting spinal motor neurons. Stabilization of mitochondrial membranes by genetic or pharmacologic manipulations also suggests a role for MOMP in neurodegeneration (table S5).

MOMP is also involved in acute toxin-induced cell death. Toxins implicated in Reye's syndrome, including salicylate, adipic, isovaleric, 3-mercaptoproprionic, 4-penenoic, and valproic acids, cause MOMP when added to purified mitochondria or to hepatocytes. Other prominent examples of toxic MOMP inducers include ethanol, CCl_4, and heavy metals or their organic derivatives (table S5).

MOMP induction is a therapeutic goal in cancer therapy. MOMP is regulated by several oncogene products, in particular the anti-apoptotic Bcl-2 family proteins, whereas several tumor suppressors induce or favor MOMP. A relative resistance to MOMP is a prominent hallmark of cancer (*34*). Experimental drugs that act on mitochondrial proteins or lipids have been shown to be therapeutic in preclinical mouse models (table S5), and some therapeutic treatments have been reported to induce MOMP, although it is not clear if direct MOMP induction accounts for their anticancer effects (table S6).

There is a great interest in developing drugs that prevent MOMP and suppress pathological cell death (table S5). Under some circumstances, MOMP is delayed or inhibited by cyclosporin A (CsA), and this can reduce the lethal effects of heavy metals or high-dose paracetamol in animal models (*35*). CsA or Bcl-2 can reduce infarct size in the heart and brain (*2*). CsA can also be used to enhance the functional recovery after hypothermic heart preservation (*36*). Several neuroprotective drugs also prevent Bax-mediated MOMP in isolated mitochondria: tauroursodeoxycholic acid probably through inhibition of Bax insertion (*37*) and dibucaine and propranolol at a later step that may involve outer membrane lipids (*38*).

Concluding Remarks

One particularly intriguing aspect that emerges from the complexity of MOMP regulation is the functional and/or physical interaction between apoptosis regulators (e.g., the Bcl-2 family) and proteins known to participate in intermediate metabolism, e.g., VDAC, hexokinase, or glucokinase. For example, the latter interacts with the pro-apoptotic protein Bad (*39*). These interactions may tie specific metabolic demands to apoptotic control and thus determine "metabolic windows" for cells to

avoid MOMP. In the absence of growth factors that regulate metabolite flow (40) or in conditions distant from optimal metabolic conditions (oxygen tension, redox potential, tissue pH, and glycolytic substrates), cells may be primed for MOMP and demise. This crosstalk between apoptosis and metabolism may contribute to the metabolic signature of cancer, the Warburg phenomenon, an increased reliance on anaerobic metabolism even in the presence of abundant oxygen. Progress at the frontiers of pathobiology will help to integrate the process of MOMP and its regulation into the physiology of the cell.

References and Notes

1. J. C. Goldstein, N. J. Waterhouse, P. Juin, G. I. Evan, D. R. Green, *Nature Cell Biol.* **2**, 156 (2000).
2. M. P. Mattson, G. Kroemer, *Trends Mol. Med.* **9**, 196 (2003).
3. J. E. Kokoszka et al., *Nature* **427**, 461 (2004).
4. L. He, J. J. Lemasters, *FEBS Lett.* **512**, 1 (2002).
5. P. C. Waldmeier, K. Zimmermann, T. Qian, M. Tintelnot-Blomley, J. J. Lemasters, *Curr. Med. Chem.* **10**, 1485 (2003).
6. D. Poncet, P. Boya, D. Metivier, N. Zamzami, G. Kroemer, *Apoptosis* **8**, 521 (2003).
7. M. C. Wei et al., *Science* **292**, 727 (2001).
8. A. Letai et al., *Cancer Cell* **2**, 183 (2002).
9. T. Kuwana et al., *Cell* **111**, 331 (2002).
10. G. Basanez et al., *J. Biol. Chem.* **277**, 49360 (2002).
11. E. H. Cheng, T. V. Sheiko, J. K. Fisher, W. J. Craigen, S. J. Korsmeyer, *Science* **301**, 513 (2003).
12. M. Deshmukh, K. Kuida, E. M. Johnson Jr., *J. Cell Biol.* **150**, 131 (2000).
13. L. K. Chang, R. E. Schmidt, E. M. J. Johnson, *J. Cell Biol.* **162**, 245 (2003).
14. M. Ott, J. D. Robertson, V. Gogvadze, B. Zhivotovsky, S. Orrenius, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 1259 (2002).
15. L. Scorrano et al., *Dev. Cell* **2**, 55 (2002).
16. O. von Ahsen et al., *J. Cell Biol.* **150**, 1027 (2000).
17. M. Karbowski, R. J. Youle, *Cell Death Differ.* **10**, 870 (2003).
18. J. E. Ricci et al., *Cell* **117**, 773 (2004).
19. J. R. Jeffers et al., *Cancer Cell* **4**, 321 (2003).
20. A. Villunger et al., *Science* **302**, 1036 (2003).
21. J. E. Chipuk et al., *Science* **303**, 1010 (2004).
22. J. I.-J. Leu, P. Dumont, M. Hafey, M. E. Murphy, D. L. George, *Nature Cell Biol.* **6**, 443 (2004).
23. M. Mihara et al., *Mol. Cell* **11**, 577 (2003).
24. M. Sawada et al., *Nature Cell Biol.* **5**, 320 (2003).
25. B. Lin et al., *Cell* **116**, 527 (2004).
26. A. Konishi et al., *Cell* **114**, 673 (2003).
27. N. Majewski, V. Nogueira, R. B. Robey, N. Hay, *Mol. Cell. Biol.* **24**, 730 (2004).
28. J. C. Rathmell et al., *Mol. Cell. Biol.* **23**, 7315 (2003).
29. L. Scorrano et al., *Science* **300**, 135 (2003).
30. M. J. Thomenius, C. W. Distelhorst, *J. Cell Sci.* **116**, 4493 (2003).
31. A. D. Badley, G. Kroemer, *Trends Pharmacol. Sci.* **24**, 298 (2003).
32. G. Mattiasson, H. Friberg, M. Hansson, E. Elmer, T. Wieloch, *J. Neurochem.* **87**, 532 (2003).
33. G. V. Putcha et al., *Neuron* **38**, 899 (2003).
34. D. R. Green, G. I. Evan, *Cancer Cell* **1**, 19 (2002).
35. D. Haouzi et al., *Apoptosis* **7**, 395 (2002).
36. K. G. Rajesh, S. Sasaguri, S. Ryoko, H. Maeda, *Transplantation* **76**, 1314 (2003).
37. C. M. Rodrigues et al., *Proc. Natl. Acad. Sci. U.S.A.* **100**, 6087 (2003).
38. B. M. Polster, G. Basanez, M. Young, M. Suzuki, G. Fiskum, *J. Neurosci.* **23**, 2735 (2003).
39. N. N. Danial et al., *Nature* **424**, 952 (2003).
40. J. C. Rathmell et al., *Mol. Cell. Biol.* **23**, 7315 (2003).
41. Supported by Agence Nationale pour le Recherche sur le SIDA, European Commission, Ligue Nationale contre le Cancer, and the French Ministry of Science (G.K.) and by NIH and Gemini Science (D.R.G.).

Supporting Online Material
www.sciencemag.org/cgi/content/full/305/5684/626/DC1
SOM Text
Tables S1 to S6
References and Notes

INTRODUCTION

From the Ignoble to the Sublime

In the 2 millennia since the first Olympic games, one principle has withstood the test of time: People are obsessed with pushing the human body to the limit. With the curtain set to rise next month in Athens on the latest Olympic Games, this Special Section goes backstage to explore some of the defining attributes of the world's greatest athletes—and their Achilles' heels.

Doping allegations against elite athletes have cast a long shadow in the run-up to this year's games. In a News report (p. 632), Vogel examines how scientific sleuths are devising new methods to unmask athletes bent on cheating their way to the top. An STKE Perspective by Handelsman (www.sciencemag.org/sciext/sports) assesses the poorly defined physiological roles of designer androgens such as THG. And in a SAGE KE News Synthesis article (www.sciencemag.org/sciext/sports), Davenport probes the promises and pitfalls of growth hormone.

What does it take to swim the fastest, throw a discus the farthest, or jump the highest? In some sports, it would seem, athletes claim such honors by birthright. Men and women from Kenya's Rift Valley dominate endurance running, for example, and runners from West Africa reign supreme as sprinters. Holden explores these apparent genetic edges in a News report (p. 637). The presence or absence of a Y chromosome creates a different kind of uneven playing field. A decade ago the best female runners were closing in on the times of their male counterparts. But Holden reports (p. 639) that the gender gap has plateaued or even increased over the past 15 years in all running events apart from the marathon.

Tending not to discriminate by gender are injuries sustained from pushing the limit. In a News report (p. 641), Stokstad describes how young gymnasts may be raising their risk of osteoarthritis and other health problems later in life. On a more positive note, information gained from studying how athletes' muscles respond to training is providing new insights on muscle growth and atrophy, the topic of an STKE Review by Sartorelli (www.sciencemag.org/sciext/sports).

Ultimately, performance comes down to mechanics. On page 643, Cho profiles Mont Hubbard, a mechanical engineer who has spent his career optimizing motion in sports. New materials can reduce physical constraints to performance. At next month's games, many of the world's best swimmers will be wearing suits with tiny ridges modeled on sharkskin that are claimed to reduce friction and drag. In a News report (p. 636), Krieger investigates those claims.

Sports scientists may not share the limelight with their study subjects, but the discipline is gathering momentum, as overview articles and profiles of early-career researchers attest on *Science*'s Next Wave (www.sciencemag.org/sciext/sports). These experts toil behind the scenes of a pursuit that, in one form or another, captivates most everyone's attention—especially every 4 years.

–RICHARD STONE

CONTENTS

NEWS

Science

A Race to the Starting Line

Scientists are scrambling to devise new methods for snaring athletes who cheat with steroids, hormones, and, someday, even extra genes

KREISCHA, GERMANY—Tucked on the wooded edge of this village in the Saxon hills south of Dresden is a drab, single-story office building with a sinister past. Until 1989 officials of the German Democratic Republic tested their athletes here to certify them as drug-free before international competitions. But it was all a charade. Many of the East German athletes, both men and women, were systematically doped up with testosterone and other anabolic steroids, often without their knowledge. It was the Kreischa lab's responsibility to ensure that the regimen was suspended long enough before a competition to flush out any traces of drugs, explains Klaus Müller, director of the Institute for Doping Analysis and Sport Biochemistry that today occupies the building. Sometimes the drug docs cut it too close. "You would hear that a certain famous athlete couldn't travel to a competition because of a 'sudden illness,' " says Müller, whose institute is part of a worldwide antidoping network. "We all knew what that meant."

That crooked chapter in German sport is over, but the practice of doping appears to be more widespread than ever. Last month world champion sprinter Kelli White received a 2-year ban from competition after admitting to having taken banned steroids and the hormone erythropoietin (EPO), which boosts red blood cell counts. Other clients of the Bay Area Laboratory Co-operative (BALCO) nutrition center in Burlingame, California, were also implicated in the scandal; as *Science* went to press, U.S. officials were investigating evidence that Olympic gold medalists Marion Jones and Chryste Gaines and world-record sprinter Tim Montgomery had been treated with banned steroids and hormones by the same lab.

In the privileged world of elite sports, avarice and the pursuit of glory continue to lead coaches and chemists astray and tempt athletes to risk health and medals. "Sport can be so magnificent and so powerful precisely because humans play the key role," says Andrew Pipe, a physician at the University of Ottawa Heart Institute and former chief medical officer to Canada's Olympic team. "It can be so depressing and sordid for exactly the same reason."

Dopers are getting better at covering their tracks, forcing researchers to invent new techniques to detect ever more subtle uses of synthetic chemicals or proteins that boost the body's ability to build muscle, shed fat, or carry oxygen. What was once the exclusive domain of analytical chemists—who searched for steroids in urine samples—now involves endocrinologists and geneticists as authorities attempt to clamp down on what could become the next illicit frontier: doping with genes for muscle

Image not available for online use.

Tainted glory. World champion Marita Koch was never directly implicated, but many of her East German teammates were given performance-enhancing drugs by their doctors and coaches.

building. "Testing gets better and better, but the opposition gets better and better too," says Don Catlin, director of the University of California, Los Angeles (UCLA), Olympic Analytical Laboratory.

Back-alley chemistry

Athletes have been seeking an artificial edge since at least the late 1800s, when runners and long-distance bicyclists used nitroglycerin and even cocaine to boost stamina and block pain. But although authorities began testing for banned substances in the 1970s, their efforts had little impact, says Peter Sonksen of St. Thomas' Hospital in London and a former member of the International Olympic Committee's (IOC's) medical commission. "For a long time there was a feeling that many sporting bodies were protecting their players," he says.

A lack of vigilance created an environment for blatant cheating. For example, a series of astounding world-record performances in the 1980s, especially in power sports such as the shot put or hammer throw, were almost certainly fueled by testosterone and other prohibited anabolic steroids, Müller says.

There is little doubt that steroids help athletes beef up. By targeting the same receptor as testosterone does, they boost the body's capacity for building muscle and erode its capacity for breaking it down. But they have manifold side effects. Although women produce some testosterone naturally, ratcheting up levels even slightly leads to increased body hair and acne and can wreak havoc with the reproductive system. In men, taking steroids suppresses natural production of testosterone, which can lead to bigger breasts, shrunken testicles, and infertility. In both sexes, high doses of the drugs damage the liver and the cardiovascular system.

As testing for steroids began to be enforced more strictly in the 1990s, use of the drugs plummeted —and the pace of record-breaking tapered off. The antidoping forces seemed to have the upper hand until 2002, when the sport world was rocked by revelations that a pair of so-called designer steroids—drugs with no legitimate medical use—had been synthesized, apparently to elude doping testers.

In one case Catlin's team detected unusually low levels of natural steroids such as testosterone, epitestosterone, and androsterone in the urine of a female cyclist, a sign that something was amiss. Probing further, his group found traces of norbolethone, an androgen developed by Wyeth in the 1960s. In animal tests, Catlin says, norbolethone appeared to be a very effective muscle builder while having relatively few masculinizing side effects. It was tested in short children and underweight patients, but Wyeth shelved the compound, apparently

Mighty Mice: Inspiration for Rogue Athletes?

The mice seem indestructible. First described in 2001, the Schwarzenegger mice, as they were dubbed by the press, have twice as much muscle as normal mice, live longer, and can recover from injuries that kill their weaker cousins. They build muscle without exercising, and they seem to defy the aging process. "As they age, they don't get weaker," says Nadia Rosenthal, a developmental geneticist at the European Molecular Biology Laboratory in Monterotondo, Italy, who created the mighty mice.

Rosenthal and her colleagues are hoping that the animals, which carry an extra copy of a gene that codes for a protein called insulin-like growth factor-1 (IGF-1), will help them understand and treat muscle-wasting diseases such as muscular dystrophy. The mice may also provide insights into wound healing and the mysterious process of regeneration. But sports authorities are worried: Can the technique that makes supermice be used in humans to create superathletes?

The practice may have already begun. Although there is no test for the molecule, a ready supply is out there. Genentech manufactured and tested IGF-1 in the 1980s but decided not to market it. Tercica, a company in South San Francisco, California, is sponsoring clinical trials of growth hormone plus IGF-1 in short children who don't respond to treatment with growth hormone alone. "If someone is doing it legally, you can bet they're doing it illegally," Rosenthal says.

But there's a subtlety to the

Wanna race? A mouse with an extra copy of IGF-1 (right) dwarfs its wild-type counterpart.

protein that should deter any athlete tempted to inject the bottled version. Some IGF-boosted mice, rather than being mighty, are in fact rather sickly. It seems that the body makes at least four forms of IGF-1. One circulates in the bloodstream and suppresses the production of human growth hormone. Another is produced in muscle tissue and appears in response to injury. Scientists are still sorting out the differences among the forms, but it's the second type that Rosenthal, H. Lee Sweeney of the University of Pennsylvania in Philadelphia, and their colleagues boosted in their mighty mice. When production of the first form is increased, mice develop oversized and weak hearts and are prone to cancer, Rosenthal says.

The form that Tercica is testing and athletes might be using is a stripped-down version of the protein, without any of the distinguishing alterations that the naturally made forms carry. It isn't yet clear how its effects compare with those of the naturally made versions, but Rosenthal says it still might be tempting to athletes: Preliminary data suggest that the Tercica version does encourage muscle growth.

What worries doping testers most, however, is the possibility that the gene-therapy technique that created the mighty mice might eventually be attempted in athletes. Such a treatment would be difficult to unmask, because the doped-up IGF-1 gene, designed to remain in muscle cells where it is produced, would not be detectable in blood or urine, says Rosenthal. Athletes would understandably be reluctant to give muscle biopsies just before a competition. That leaves yet another conundrum for testers to resolve in the fight against doping.

–G.V.

because of toxic side effects. Evidently, someone was cooking up a new supply.

A whistleblower made the second discovery possible. In June 2003 Catlin received the residue from a used but empty syringe from the U.S. Anti-Doping Agency. A track coach had sent it to the authorities, suggesting that they take a careful look. Within a few weeks, Catlin and his colleagues had identified tetrahydrogestrinone (THG). The new chemical, which had never before been described, resembles two steroids banned for use by professional athletes: gestrinone, prescribed occasionally for the treatment of endometriosis, and trenbolone, which has some uses in veterinary medicine. Both steroids have powerful anabolic effects, and the UCLA team quickly suspected that the derivative had been designed to activate the same receptors while foiling standard screens for known steroids. When authorities tested urine stored from previous competitions, they found at least a dozen THG-tainted samples, many from athletes who had connections to BALCO.

Because routine screening would never

have caught THG, doping testers were confronted with the prospect of having to develop ways to detect an incalculable array of steroids and other chemicals that might play a similar performance-enhancing role. "The THG story tells us very convincingly that there are people out there who are scheming to develop new entities to give to athletes," says Catlin. "We've studied the chemistry, and there's essentially no end to the possibilities. Are there others out there? There certainly are." They just haven't been identified yet, he adds.

Some labs are hoping to defeat dopers at their own game. Wilhelm Schänzer and his colleagues at the Institute of Biochemistry of the German Sport University Cologne have begun churning out more than a dozen potential designer agents by tinkering with existing steroids. "We're trying to think in the same way as those who are trying to make new compounds," Schänzer says. His group uses mass spectrometry to profile the concoctions and identify signals that might betray illicit compounds in bodily fluids.

THG presented a legal challenge as well. Lawyers for athletes who tested posi-

tive argued that the authorities couldn't demonstrate that the substance is an anabolic steroid, and therefore it could not be classified as a banned substance. Indeed, the chemical's effects in animals—much less humans—had never been characterized in a legitimate lab; standard animal tests take many months. Under court-imposed time constraints, scientists resorted to a quicker solution, a test originally designed to ferret out environmental pollutants that mimic hormones. The test uses yeast cells altered to make the human version of the testosterone receptor as well as a luminescent protein that glows when the receptor is activated. Using the test, David Handelsman of the ANZAC Research Institute in Concord, Australia, found that THG lights up the cells more brightly than standard anabolic steroids such as trenbolone and even testosterone.

The confirmation came just in time to support the case against European champion sprinter Dwain Chambers, who had tested positive for THG in August 2003. (Chambers has said that he ingested the compound unknowingly in a supplement

provided by BALCO.) In February, U.K. Athletics banned him from running in competitions for 2 years. Chambers had been considered a favorite for the gold medal this summer in Athens, but according to British Olympic Association rules, he is banned from the Athens games.

The bioassays may soon join a growing arsenal that scientists are assembling to thwart the use of new designer steroids, says Handelsman. He and his colleagues, for example, are working on a simple test to compare the amount of testosterone normally present in the urine of men and women with the total steroid load, as measured by the bioassays. "If there's a gap, then that suggests there's an unidentified substance there," Handelsman says.

The workhorse of steroid detection, the mass spectrometer, could also be put to innovative use. Even if an analysis fails to flag unexpected side chains or telltale peaks, it can reveal subtle differences in the ratio of carbon isotopes that can help identify the origin of organic molecules. An unusual ratio of carbon-12 to carbon-13 in certain molecules can raise a red flag in a doping test. If a steroid molecule has a ratio typical of a plant rather than an animal, it is a sign that it comes from an outside source, says Schänzer.

In pursuit of oxygen

Unknown steroids are hard enough to pin down; injections of naturally occurring hormones are even more elusive. Hormone levels fluctuate from hour to hour and from person to person, so measuring absolute amounts can't nail a doper. To do that, scientists must find secondary signals indicating that the body's normal chemistry has been tampered with.

For years, some athletes took advantage of the dearth of detection methods to pump themselves up with EPO. The hormone, produced mainly in the kidneys, stimulates the body's production of red blood cells so that the blood carries more oxygen. People living at high altitudes produce more EPO naturally to compensate for the lower oxygen concentration in the air. Athletes often take advantage of that trick, training at high altitudes for competitions held nearer sea level. But when recombinant EPO, used to treat anemia, became available in the late 1980s, it spawned a doping epidemic.

The practice is dangerous. If blood has too many red blood cells, it can become too viscous for the heart to pump effectively. EPO is thought to have played a role in the deaths of more than a dozen

Dutch and Belgian cyclists who died of sudden heart attacks in the 1980s, just after EPO became available in Europe. Despite the risks, EPO's use was apparently widespread in the 1990s as scientists raced to figure out how to detect its use.

The first EPO tests, introduced a decade ago, set a limit for hematocrit, the percentage of red blood cells in the blood. But that test is flawed, as it cannot tell whether an athlete has

Banned. Dwain Chambers, who tested positive for the steroid THG, is barred from the British Olympic team.

Take a deep breath. Blood samples will be collected from athletes in Athens to check for erythropoietin and a range of other substances.

used EPO to boost his or her hematocrit to a level just below the allowed limit.

In 2000, in time for the Olympic Games in Sydney, Australia, the IOC introduced a combined blood and urine test for EPO. The blood test measures, among other things, the concentration of hemoglobin and the level of reticulocytes—immature red blood cells—in the blood. Testers look for unusually high levels or sudden changes from previous tests to tip them off to possible dopers. The test has one major advantage: It can detect signs of EPO use weeks after an athlete takes it. But because it does not measure illegal EPO directly, it cannot prove a doping allegation.

A second method allows testers to spot traces of recombinant EPO directly in urine. Because the recombinant version is produced in animal cells, it carries slightly different sugars in its side chains than the natural version. These differences show up in electrophoresis, which measures the distance proteins chug through a gel under the influence of electricity. The concentration of EPO in urine is fairly low, however, so the test could be foiled if an athlete takes diuretics or other urine-increasing drugs.

The bottom line is that the current tests simply don't cut it. "Athletes are getting around the EPO tests all the time," Catlin says. Officials of the World Anti-Doping Agency (WADA) agree. "We need cheaper and more sensitive tests for EPO," says Olivier Rabin, WADA's scientific director.

WADA is also funding projects to tackle an old-fashioned doping technique that the organization claims is back in vogue since the introduction of EPO tests. Called blood doping, it involves an athlete either receiving blood transfusions—enriched in red blood cells—from donors, or removing an athlete's own blood, spinning it to concentrate the red blood cells, then reinfusing it right before competition. Although the techniques don't involve foreign chemicals, they are banned by sports organizations on safety grounds.

A growing threat

One of the compounds that BALCO clients are accused of abusing is something that doesn't show up in any standard doping tests: human growth hormone (hGH). The protein is part of a biochemical cascade that spurs muscle buildup and the shedding of fat. It's used legitimately to

treat children who lack the protein and are unusually short. But like EPO and legitimate steroids, it too has been hijacked for use in athletes. Although its effects in healthy athletes are unclear, doping experts suspect that its use is widespread—especially because authorities have not yet introduced an official test for the compound.

That's a high priority, however, and scientists say they have several tests ready for the Athens Games. WADA officials are circumspect about whether they will use any of the tests for hGH in August. "Athletes know it is on the banned substances list," says Rabin, and should expect to be tested.

Detecting hGH is even harder than detecting EPO, because it doesn't have telltale sugars to betray artificial versions. But in a lucky break for doping sleuths, the pituitary gland's production of growth hormone is rather messy. The gland makes a mixture of variations of the protein as well as protein fragments. The manufactured version, on the other hand, is much cleaner, consisting chiefly of one of the heavier versions, so when someone shoots up with the recombinant protein, the ratio of the different forms is skewed. Endocrinologist Christian Strasburger of the Charité University Clinics in Berlin and his colleagues at the Medizinische Klinik Innenstadt at the University of Munich have developed an immunoassay that measures the ratio of the two forms. The test seems extremely reliable, Strasburger says.

Another group led by Sonksen of St. Thomas' Hospital has developed a method to measure the effects of growth hormone on the production of other proteins, including insulin-like growth factor-1 (IGF-1) and collagen. The test is not as clear-cut as that developed by Strasburger and his colleagues, but it can detect the effects of hGH weeks after someone has injected it. The Strasburger method works best 24 to 36 hours after injection.

Those who go to the trouble and expense—a month's dose costs more than $2500—may not be getting their money's worth. "It looks as though growth hormone is fool's gold," says Ken Ho, an endocrinologist at the Garvan Institute of Medical Research near Sydney. "In the normal person with normal levels of growth hormone, adding extra has not been shown to confer a benefit." Yet, Ho says, "at the end of the day, if a 0.01% advantage is the difference between winning and losing," a minuscule boost from growth hormone—even if it's purely psychological—might help an athlete to victory.

Tackling this murky question, Ho's group is giving growth hormone to healthy volunteers both to screen for biochemical

changes that might be picked up in a doping test and to look for performance-enhancing effects. Whether the benefit is real or not, the hormone is on the list of banned substances, and athletes caught using it will forfeit any medals they receive next month in Athens.

Self-assembled superathletes

In a case that made headlines this summer, doctors described a young boy in Berlin who seems destined for athletic greatness. The boy was born with a mutation that turns off the gene for myostatin, which in animals seems to block the activation of muscle stem cells. Mice and cattle that carry myostatin mutations have twice as much muscle as normal animals. At 4-and-a-half years old, the boy had the physique of a mini-bodybuilder and could hold out two 3-kg dumbbells with his arms extended, his doctors reported in the 24 June *New England Journal of Medicine*. Some experts are thrilled: They suggest that the mutation could be exploited as a treatment for muscle-wasting diseases.

But antidoping officials are cringing. They fear that gene therapy could soon be the next fad among athletes. Their nightmare scenario is athletes injecting a retrovirus bearing a myostatin-blocking gene or another muscle-building gene such as IGF-1 (see sidebar on p. 633). Once the gene is incorporated into cells, it begins to pump out its products and build muscle, but the illicit source would be extremely difficult to trace.

Clinical trials of gene therapy for fatal diseases have been fraught with problems, including the death of one volunteer and the development of leukemia in other patients. But that might not stop some athletes. "These [gene therapy] methods remain extremely risky," says Rabin. "But on the other hand we know that some athletes are willing to take incredible risks. THG went straight from the test tube to the athletes, with no proper testing."

WADA is funding efforts to detect gene doping, either through traces of retrovirus vectors or by spotting indirect effects of gene boosting, Rabin says. Almost any gene doping would influence a wide range of

other genes, and changes in those might be traceable with ever more sensitive tests that can flag gene expression, he says.

As doping sprouts more Medusa-like heads, authorities may be forced to develop personalized tests. Ideally, Rabin says, each athlete would submit a biological "passport" containing highlights of their blood chemistry. "If we then saw an abnormal change, we could follow it up," he says. Microarrays that measure the expression levels of thousands of genes at once could betray blips that might result from gene doping. Ideally, he says, "in a single drop of blood, we'll be able to detect changes based on any genes that are modified." For now, though, such tests would cost thousands of dollars per athlete, prohibitively expensive for most testing organizations.

Müller doesn't foresee a breakthrough in antidoping efforts anytime soon. His experi-

Natural boost. A Berlin child who carries a mutation in the myostatin gene has had bodybuilder muscles since birth.

ence as a medical scientist in East Germany, where athletes were lavishly funded as an international propaganda tool, left him questioning the value of top-level sports. "We are not dealing here with problems of human existence or survival," he says. "The world will not come to an end if dopers go uncaught." His motivation, he says, is the example that elite athletes set for millions of amateurs: "It's important for people to be able to understand that you can do amazing things without doping."

Unfortunately, many elite athletes don't buy that message. "We're never going to eliminate [doping] completely," says the University of Ottawa's Pipe. The contest will continue, with both sides intent on raising their game to the next level.

–GRETCHEN VOGEL

SPECIAL SECTION

NEWS

Do Pool Sharks Swim Faster?

New swimsuits with tiny ridges modeled on sharkskin are all the rage. Experts are split, though, about whether the high-tech suits reduce drag

When the International Olympic Committee declared at the last minute that "sharkskin" swimsuits were legal for the 2000 games in Sydney, Australia, swimmers had just 1 day to decide whether to ditch their tried-and-true swimwear for the flashy new ridged bodysuits that could potentially shave seconds off their times. In a traditionally low-tech sport, choice of equipment suddenly took on make-or-break significance.

For some of the world's best swimmers 4 years later, the choice is no longer theirs. Stars such as Michael Phelps, Inge de Bruijn, Lindsay Benko, and Grant Hackett are obligated to wear Fastskin, a swimsuit made by Speedo, their sponsor. The companies TYR and Arena also have swimmers wearing their versions of the suit, inspired by sharkskin and intended to cut down on drag, the friction that slows any body moving through water. On its Web site, Speedo states that Fastskin reduces drag up to 4%. Other companies make similar claims.

But whether the suits work is hotly debated. Critics dispute the explanations of how the suits reduce drag. "The claims that the manufacturers make about the science behind what these suits do are outrageously ignorant," asserts Brent Rushall, a professor of exercise and nutrition at San Diego State University in California.

The concept itself is not unreasonable. Sharkskin is textured with ridges, known as "riblets," that reduce the amount of skin surface area that comes in contact with water. That allows sharks to glide through water with much less friction than one would expect. In 1987, the hull of an America's Cup sailboat was textured to mimic the skin of pelagic sharks; it won the race and was so fast that the texturing was outlawed. The advantage made sense: Pelagic sharks cruise long distances at speeds approaching those of sailboats, around 15 knots, or 28 kilometers per hour, so the benefit from the riblets was transferred from beast to boat.

Manufacturers say that the sharkskin suits are grounded in sound science. The suits are designed to lower two types of drag: skin friction and pressure drag. The trick for reducing skin friction is making the riblets just the right size. If they are too big or too small, water floods the valleys, creating more surface area than a standard suit has, says Barry Bixler, an aerospace engineer who models fluid dynamics for Speedo. Pressure drag, meanwhile, occurs when water flowing over irregular shapes, such as eyebrows, chins, and breasts, peels off in sheets, creating suction. Imagine water flowing over a beach ball until it reaches a critical angle and falls away, creating a vacuum that sucks at the ball. To reduce this effect, the Speedo suit is dimpled, like a golf ball, in strategic places. The indentations set up miniature eddies; turbulent water is more likely to stick to the swimmer than is smooth-flowing water. The suits are designed to keep "sticky" water continuously flowing over the body, thus reducing pressure drag.

According to Bixler, Speedo's models are based on body scans of two Olympic medalists, a man and a woman (Speedo will not reveal their names), in a streamlined diving position. Mannequins of the swimmers match the model's calculated drag predictions to within a few percent, Bixler says. However, neither the mannequins nor the computer-generated swimmers, the basis for calculations of passive drag, stroke or kick the way a human would. "It's difficult to model that," says Bixler. However, he says, "we know that if passive drag is less, active drag will also be less." By Speedo's calculations, the bodysuits reduce passive drag 4% in men and 3% in women, whose hips and breasts tend to increase pressure drag.

One skeptic is Huub Toussaint, a biomechanist at the Free University in Amsterdam. The manufacturers "take the swimmer and put him underwater in a flow tank with [a] steady flow, … and they ignore that in reality, the swimmer is moving." In *Sport Biomechanics* in 2003, Toussaint and colleagues compared swimmers propelling through a pool wearing training suits with swimmers wearing sharkskin suits. The study found no significant reduction in drag from the sharkskin suits.

Rushall, meanwhile, argues that the flow of water over a boat's hull is nothing like that over a swimmer. The fastest humans thrash through water at a maximum speed of about 7 kilometers per hour, a quarter of the speed of a pelagic shark or a sailboat. "Whatever animal [the swimsuit manufacturers] are trying to reflect is not going to be swimming as slowly as a human," says Rushall. A texture that lessens friction at shark speeds may even increase friction at human speeds, he says. Rushall claims that a Japanese swimsuit maker, which he declined to name, had tested a sharkskin-style surface material and found that it had reduced friction only at very specific angles of entry to the water. The company concluded that any reduction of drag on the irregular bumps and indentations of the human body would be minimal and dropped the project.

Independent tests are hard to carry out. Bixler confirms that the suits reduce both friction and pressure drag but says that Speedo has barred him from publishing his modeling data on the suits. Speedo and other manufacturers contacted by *Science* say that the data are trade secrets.

The uncertainty has led some to focus on the test that really matters: Olympic times. As a result of improvements in training, nutrition, and technique, swimmers steadily improve in a statistically predictable way, says Joel Stager, a physiologist at Indiana University, Blooming-

ton. The data wobble when performance enhancers such as doping or technological breakthroughs enter the picture. After analyzing swim times in the Sydney Olympics, Stager says, "If the suits worked, we shouldn't be able to predict performance. But lo and behold, we were pretty much dead-on except for two events." Of those two, the men's 100-meter breast stroke was faster than predicted, with all the athletes wearing conventional swimsuits. The other event, the women's 200-meter backstroke, was slower than predicted—with everyone wearing the sharkskin suits.
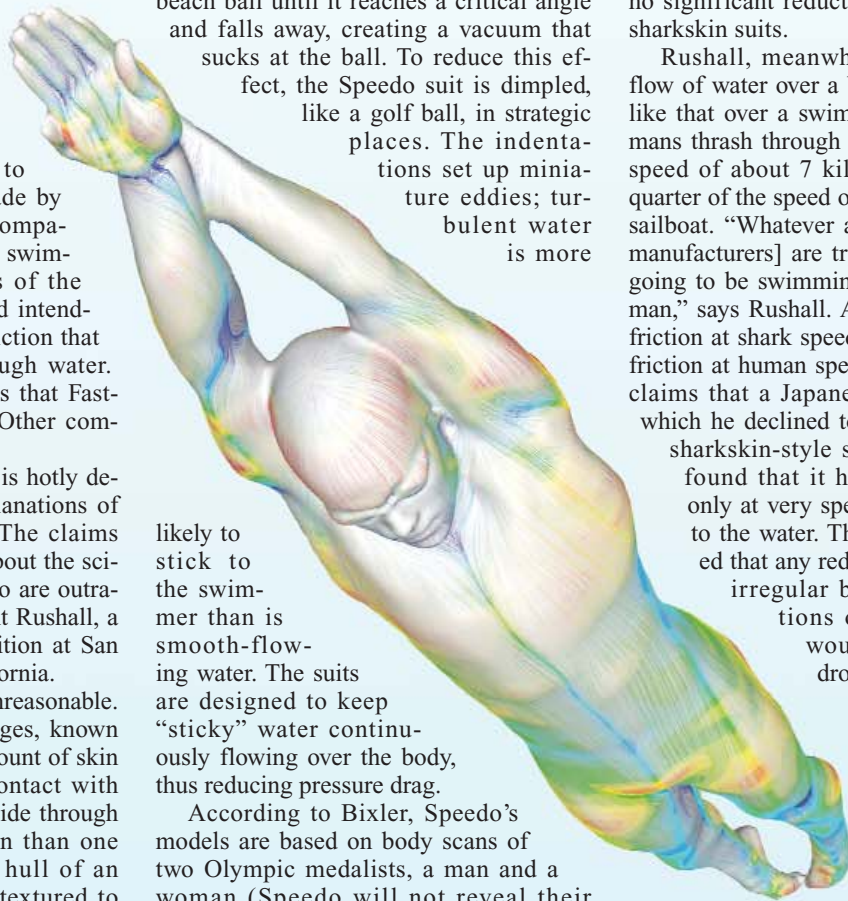
Next month's Olympics in Athens will provide a high-profile venue for the sharkskin suits to sink or swim. Manufacturers have had 4 years to hone their product, and more athletes will have practiced with them and have their own specially fitted versions.

Effective or not, the suits are catching on. "I haven't seen a freestyler race in anything but a Fastskin for as long as I can remember," says Kevin Swander, the 2004 Big Ten champion, who competed in the Olympic trials in Long Beach, California, in July. Swander likes the bodysuits and has worn one at every single meet of his college career. "I don't know if it reduces drag. I like the way it feels," he says. "The biggest thing with these suits is the mental aspect. The suit supposedly makes you swim faster, so [when] you're wearing them—you swim faster."

Stager, for his part, is wistful for the days when anyone with a $30 swimsuit could compete. When modern Olympians put their full-body sharkskin suits on, "it's like they become invisible," he says. "They lose their identity." One thing has not changed, at least: In the pool, speed always trumps appearance.   –KIM KRIEGER

Peering Under the Hood of Africa's Runners

Kenyans dominate endurance running, and West Africans excel as sprinters. With a physiological explanation in hand, researchers are now probing the genetics of this geographic mastery

In 1968, a Kenyan runner named Kip Keino emerged as a shining star of the Mexico City summer Olympics, setting a world record in the 1500-meter race. Year after year Keino's success has been followed by equally dazzling feats by his compatriots: Kenyan men now hold world records in the 3000-meter track race, the 15-, 20-, and 25-kilometer road races, the half-marathon, and the marathon. Kenyan men have won 13 of the last 14 Boston marathons. Kenyan women are also rising fast: They hold half of the top 10 marathon times and world records in 20-, 25-, and 30-km track races. What is even more remarkable is that most of these athletes come from a small area in Kenya's Rift Valley, from a group of tribes called the Kalenjin who number little more than 3 million people.

Theories abound about what Kenya-born writer and runner John Manners calls "the greatest geographical concentration of achievement in the annals of sport." Is it the high altitude that fosters big lungs and efficient oxygen use? Is it their maize-based diet? Or the fact that many children run to school? A grueling training regimen, perhaps?

Such questions have inspired a handful of researchers to try to define the Kenyan magic. Meanwhile, scientists are unraveling why athletes whose ancestors come from the other side of the continent—West Africa—have emerged as the world's fastest sprinters.

Fuel economy

Leading the charge in penetrating the Kenyan mystique has been Bengt Saltin, a Swedish physiologist who heads the Copenhagen Muscle Research Centre in Denmark. In the 1990s, Saltin's group began comparing Kenyan and Scandinavian runners by scrutinizing their physiological makeups and assessing the "trainability" of novice runners in both countries.

A decade later, the scientists have ruled

Running revolutionary. Kenya's Kip Keino in 1972.

out most of the popular explanations for Kenyans' domination of running. Altitude is not the key to the riddle, they have found, because there's no difference between Kenyans and Scandinavians in their capacity to consume oxygen. And the Kenyan diet is on the low side for essential amino acids and some vitamins as well as fat, says Dirk Christensen of the Copenhagen center: "In spite of the diet, they perform at high level." The running-to-school hypothesis was demolished as well: Kenyan children aren't any more physically active than their Danish peers. Do Kenyans try harder? The researchers found that the Danes actually pushed themselves harder on a treadmill test, reaching higher maximum heart rates.

An important clue is the ability of Kenyans to resist fatigue longer. Lactate, generated by tired, oxygen-deprived muscles, accumulates more slowly in their blood. Comparisons of lactate levels have suggested to Saltin's group that Kenyan runners squeeze about 10% more mileage from the same oxygen intake than Europeans can.

Just as more aerodynamic cars get better gas mileage, the Kenyan build helps explain their fuel efficiency. A recent British TV documentary described the Kalenjin as possessing "birdlike legs, very long levers that are very, very thin [on which they] bounce and skip" along.

Saltin's group has quantified this observation. Compared with Danes, the thinner calves of Kenyans have, on average, 400 grams less flesh in each lower leg. The farther a weight is from the center of gravity, the more energy it takes to move it. Fifty grams added

to the ankle will increase oxygen consumption by 1%, Saltin's team calculates. For the Kenyans, that translates into an 8% energy savings to run a kilometer. "We have solved the main problem," declares Henrik Larsen of the Copenhagen center. "Kenyans are more efficient because it takes less energy to swing their limbs." Other scientists say the jury is still out on the Kenyan question. But "I think Saltin is probably the most correct that anyone is at the moment," says physiologist Kathryn Myburgh of the University of Stellenbosch in South Africa, who is exploring the role of Kenyans' training.

However, slim lower legs are not the whole story. Kenyan runners also have a higher concentration of an enzyme in skeletal muscle that spurs high lactate turnover and low lactate production. Saltin says that this results in an "extraordinarily high" capacity for fatty acid oxidation, which helps wring more energy out of the muscles' biochemical reactions. Because intense training alters the body's biochemistry, Saltin says that he can't say for sure whether the ezyme levels are due to genes or training. But he adds, "I think it's genetic."

Research in South Africa jibes with the Copenhagen group's findings. A team led by exercise physiologist Adele Weston of the University of Sydney, Australia, compared black South Africans, whose running strengths are similar to those of Kenyans, with white runners. The two groups had similar VO2 max values—that is, when putting out maximum effort, they used up the same amount of oxygen per kilogram of body weight per minute. But the black runners were more efficient in their oxygen consumption, lasting on a treadmill at maximum speed for twice as long as the whites. As with the Kenyans, the black South African runners accumulated less lactate and had higher levels of key muscle enzymes.

A little more twitchy

Whereas East Africans dominate long-distance running, West Africans have surged

Triumph of fast twitch. Carl Lewis, a U.S. sprinter with West African roots, winning the 400-meter race at the 1992 Olympics in Barcelona.

to the fore in short-distance events. Little research has been done on West Africans, but there's powerful circumstantial evidence for some physical advantages, as presented by Jon Entine in his book *Taboo: Why Black Athletes Dominate Sports and Why We're Afraid to Talk About It.* Athletes of primarily West African descent—which includes the majority of U.S. blacks—hold all but six of the 500 best times in the 100-meter race, "the purest measure

Out of East Africa. In less than 2 decades, Kenyans came to dominate the top 20 performances in six races ranging from 800 meters to the marathon.

1986

48.3 · 12.5 · 13.3 · 3.3 · 13.3 · 4.2 · 5.0

2003

55.8 · 0.8 · 11.7 · 29.2 · 2.5

- Asia
- Kenya
- South America
- Africa excl. Kenya (includes U.S. blacks)
- North America
- Europe
- Australia

of running speed," says Entine, whose book set off a broad debate on the subject.

Various studies have shown that West African athletes have denser bones, less body fat, narrower hips, thicker thighs, longer legs, and lighter calves than whites. But the differences between East and West Africans are even more striking. The fabled Kenyan runners are small, thin, and tend to weigh between 50 and 60 kilograms, whereas West African athletes are taller and a good 30 kilograms heavier, says Timothy Noakes, a prominent exercise physiologist and researcher at the University of Cape Town.

The differences don't stop with body shape; there is also evidence of a difference in the types of muscle fibers that predominate. Scientists have divided skeletal muscles into two basic groups depending on their contractile speed: type I, or slow-twitch muscles, and type II, fast-twitch muscles. There are two kinds of the latter: type IIa, intermediate between fast and slow; and type IIb, which are superfast-twitch. Endurance runners tend to have mostly type I fibers, which have denser capillary networks and are packed with more mitochondria. Sprinters, on the other hand, have mostly type II fibers, which hold lots of sugar as well as enzymes that burn fuel in the absence of oxygen. In the 1980s, Claude Bouchard's team at Quebec's Laval University took needle biopsies from the thigh muscles of white French Canadian and black West African students. They found that the Africans averaged significantly more fast-twitch muscle fibers—67.5%—than the French Canadians, who averaged 59%. Endurance runners have up to 90% or more slow-twitch fibers, Saltin reports.

Bouchard, now at Louisiana State University in Baton Rouge, says his team looked at two enzymes that are markers for oxidative metabolism and found higher activity of both in the West Africans, meaning they could generate more ATP, the energy currency of the cell, in the absence of oxygen. The study suggests that in West Africa there may be a larger pool

Road test. Masks monitored Kenyans' oxygen use.

of people "with elevated levels of what it takes to perform anaerobically at very high power output," says Bouchard.

Although training can transform super-fast-twitch type IIb fibers into the hybrid type IIa, it is unlikely to cause slow- and fast-twitch fibers to exchange identities. Myburgh says there is evidence that, with extremely intensive long-distance training, fast IIa fibers can change to slow type I fibers. So far, however, there is no evidence that slow-twitch fibers can be turned into fast-twitch ones. As an athlete puts on muscle mass through training, new fibers are not created, but existing fibers become bigger.
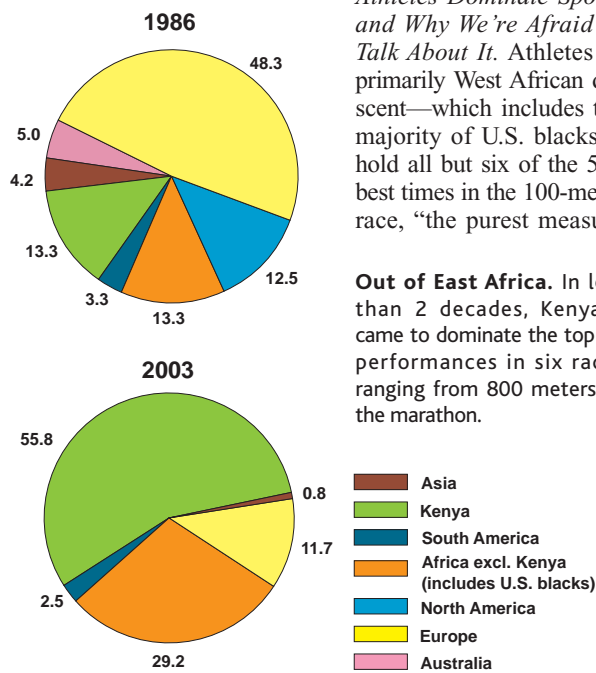
Running ACEs

The differences in physique and muscle makeup that underlie the dominance of Kenyan endurance runners and West African sprinters doubtless have a strong genetic component. But researchers are only just getting off the starting mark in the search for genes that influence running performance. Bouchard's group, for example, is collecting DNA samples from 400 runners and other top endurance athletes from the United States and Europe, but he says they haven't spotted any running genes yet.

There are a couple of intriguing possibilities, though. In 1999, a team headed by Kathryn North of the Children's Hospital at Westmead in Australia described two versions of a gene that affects production of α-actinin-3, a protein found only in fast-twitch muscles. They found the less efficient version of the gene—which results in poorer energy conversion—in 18% of the members of a group of Caucasians. In 2003, North's group reported in the *American Journal of Human Genetics* that only 6% of a group of sprinters had the gene defect; 26% of endurance runners had it. The authors surmise that α-actinin-3 helps muscles generate "forceful contractions at high velocity."

Alejandro Lucia Mulas of the European University in Madrid is taking DNA samples from Eritrean runners to explore another candidate: different versions of the gene for angiotensin-converting enzyme (ACE). Lucia says the less active version, or I allele, of this gene is associated with less muscle, less fluid retention, and more relaxed blood vessels—which would enhance oxygen uptake—and appears to be more prevalent in endurance runners.

And in Scotland, sports physiologist Yannis Pitsiladis has launched a major onslaught on the Kenyans' secrets with the

International Centre for East African Running Science. Headquartered at the University of Glasgow, the virtual center will bring together research on demography, diet, and socioeconomic factors as well as

genes. Pitsiladis says he has spent the last 3 years in East Africa collecting DNA samples from their "living legends" and now has DNA from 404 Kenyan and 113 Ethiopian athletes. His team has found a higher prevalence of the I allele for the ACE enzyme in male marathoners compared with men from the general Ethiopian population. But Pitsiladis thinks his numbers may lack significance given the variability of the trait in African populations. "At the moment there is no evidence" that East Africans have a genetic advantage in running, he says.

None of the data negate the importance of cultural habits and training. But as Entine quotes anthropologist and sports science expert Robert Malina, who is retired from Michigan State University, "Differences among athletes of elite caliber are so small that if you have an advantage that might be genetically based … it might be very, very significant."

Next month's Olympic games in Athens should demonstrate yet again that West African runners are built for speed and Kenyans built to endure.

–CONSTANCE HOLDEN

NEWS

An Everlasting Gender Gap?

For a while, female runners were closing in on their male counterparts. Now they're barely keeping the guys' taillights in sight

When the U.K.'s Paula Radcliffe ran the London Marathon last year in just under 135 minutes, shaving almost 2 minutes off the record she set in 2002, the whispers started again: Are women improving their performance so quickly that one day they may compete on the same tracks with men?

Expert opinion suggests that day will remain elusive—as long as women retain female bodies. The gap between the sexes appears to have plateaued, with women performing at about 90% of male levels. Apart from the marathon, "world records for women have been absolutely static" for more than a decade, notes Kenya-born journalist and running expert John Manners.

That plateau wasn't evident 12 years ago. In a letter to *Nature* published on 2 January 1992, titled, provocatively, "Will women soon outrun men?," Brian L. Whipp and Susan Ward of the University of California, Los Angeles, looked at the world records of five standard Olympic running events, from the 200-meter dash to the 26-mile (42-kilometer) marathon, from the 1920s through 1990. They found that

women were improving their times at double the rate of men in the short distances and were narrowing the gap even faster in the marathon, in which record-keeping for women only started in 1955. "The gap is progressively closing," the authors wrote. At that rate, they projected that marathon times could converge by 1998 and that gender differences in all races could disappear by 2050. In 1996, a poll by *U.S. News and World Report* reported that two-thirds of Americans believed that "the day is coming when top female athletes will beat top males."

Absent unforeseen genetic or hormonal interventions, men, it seems, will maintain an advantage. That's due largely to their steady supply of a performance drug that will never be banned: endogenous testosterone, which boosts muscle power and oxygen capacity. The typical young man has a maximum oxygen use capacity, or VO2 max, of about 3.5 liters per minute, compared with 2 liters for a woman, says physiologist Stephen Seiler of the Institute of Health and Sport at Agder College in Kristiansand, Norway. Although individual

CREDIT: THOMAS NOLAN

Percentage of all track-and-field medals won by Eastern bloc athletes.
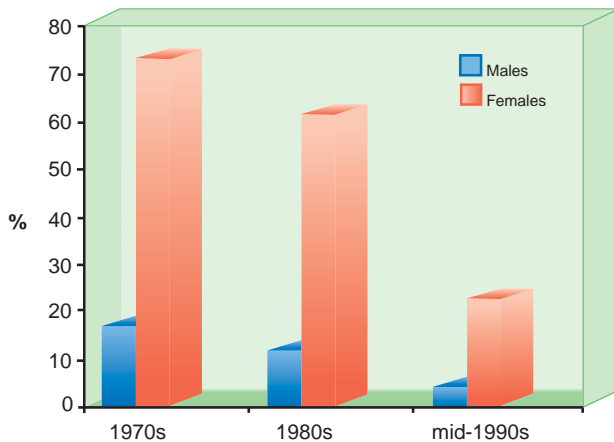
Fall of eastern superwomen. Drug use helped inflate female performance before the fall of communism.

levels of testosterone vary widely, males tend to have at least 10 times as much of the stuff as women. The hormone stimulates the creation of red blood cells, which means that men's blood holds about 10% more of the oxygen-carrying protein hemoglobin. But oxygen is not the whole story. Kirk Cureton of the University of Georgia School of Health and Human Performance in Athens compared the performance of male and female athletes on an exercise bicycle after scientists had withdrawn blood, leaving the subjects with equal amounts of hemoglobin in circulation. That reduced but did not eliminate the sex difference in VO2 max, indicating that other factors, particularly musculature, play into the difference.

Men have more muscle and larger hearts in relation to body size, says Dirk Christensen, an exercise physiologist at the University of Copenhagen. This affects aerobic capacity: He says that a trained woman's heart can pump out the same volume of blood as a man's can, but it has to work much harder to do so.

Because testosterone spurs growth of muscle tissue, it also affects anaerobic capacity—the ability to produce energy quickly without oxygen—which gives males an edge in sprinting as well. The primary energy for the intense bursts of power required for sprints is generated anaerobically, explains retired Michigan State University anthropologist and sports expert Robert Malina. Indeed, after launching themselves for a 10-second sprint, some athletes "don't take another breath till it's all over," he says. (Endurance running in contrast relies almost exclusively on aerobic energy.) More muscle means more of the two main anaerobic energy sources: phosphocrea-

tine and glucose.

Recent records support the gender-gap plateau, Seiler says. A few years ago, he and writer Steve Sailor analyzed results from Olympic games and world championships of the International Association of Athletics Federations between 1952 and 1996, selecting events in which men and women ran under the same conditions. They found that if the marathon—which wasn't an Olympic event for women until 1984—were excluded, the mean performance gap for running events increased from 11% in the mid-'80s to 12% in the mid-'90s. They also observed that men's world records were broken far more often in the '90s than women's—largely due to the extraordinary performance of East African runners (see p. 637).

Seiler recently updated these numbers for *Science*. He reports that from the world records in the eight main running events from 100 meters to the marathon, seven suggest an increasing gender gap. The

Runaway record holder. Paula Radcliffe keeps carving minutes off the marathon.

marathon is the exception: The gap has narrowed from 11.9% to 8.4%, thanks to Radcliffe's new record. Seiler says the current average gap is now 11.01%, up from 10.4% in 1989. In short, he says, "at the highest levels of performance, the gender gap in running performance has actually widened over the last 20 years."

Much of the female record is clouded by drug use, especially the records set in the 1970s and '80s by Eastern European women that have never been bested. In 1984, 38 women, mostly from the East Bloc, ran 1500 meters in under 4.05 minutes, according to Jon Entine in his book *Taboo*. In 1991, only nine did.

Although the impressive gains in female marathon performance have suggested to some observers that women have greater endurance than men, physiologist Henrik Larsen of the Copenhagen Muscle Research Centre says that's not so: "Women had not developed long distance; that's why the improvement is much greater on the marathon. We don't see any higher oxidative capacity in women." Exercise physiologist Timothy Noakes of the University of Cape Town, South Africa, agrees. A smaller body frame gives women an edge on endurance, he says, but men can run 10% faster even when the difference in body size is controlled for.

Whipp says he's still keeping an open mind on the subject of male-female competition. He told *Science* that he and colleagues are currently working to extend their analysis of world running records to 2003. His team has looked for "a pattern of response that would suggest a physiological limit," he says, but so far has found none. "There is no evidence at the beginning of the 21st century that the human athlete has reached the limit of [his or her] potential," Whipp says.

But that appears to be a minority view. "We *are* approaching the limits of human performance in a lot of the one-dimensional events like the 100-meter sprint or marathon," says Seiler. "Records will continue to be broken, but the price is extremely high. And the percentage of the population that has the genetic potential to excel at this level is infinitesimal." As for the gender gap in running, he defers to Norway's marathon queen Grete Waitz, setter of world records in the 1970s and '80s, who said: "As long as women are women, I don't think they will surpass men."
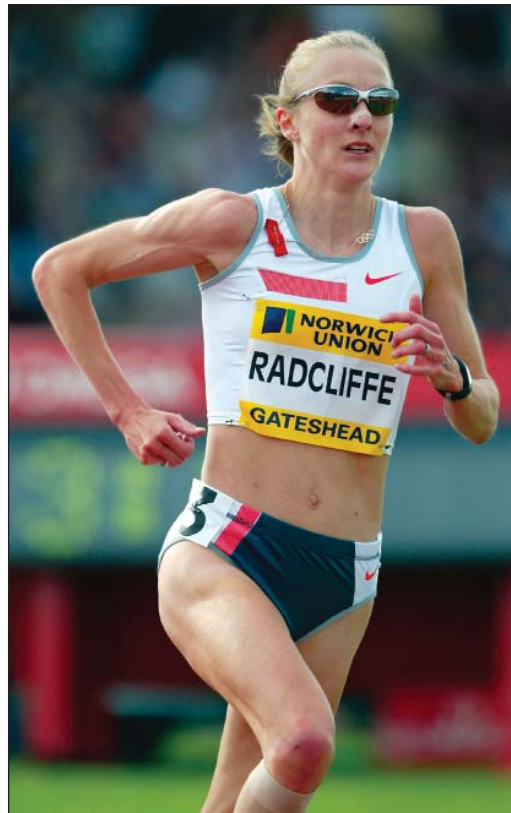
—CONSTANCE HOLDEN

NEWS

Graceful, Beautiful, and Perilous

As gymnastics routines grow ever trickier, experts worry that children are being pushed beyond their limits—and are paying with their health

When Natalia Yurchenko introduced a new vault at the 1983 World Championships in Budapest, it helped her win a gold medal. The move is fiendishly difficult: Yurchenko would sprint 20 meters, cartwheel, land backward on a springboard, and launch herself into the air. Arching her back, she would reach for a padded apparatus called a horse and then propel herself into the air again, somersault one-and-a-half times, twist, and land facing backward.

Following that dizzying lead, gymnasts set out to conquer the Yurchenko vault—and injuries mounted. "Every single country had problems," recalls William Sands, head of sport biomechanics and engineering at the U.S. Olympic Training Center in Colorado Springs, Colorado. The vault became notorious in 1988, when 15-year-old Julissa Gómez broke her neck while attempting it at the World Sports Fair in Japan. She was paralyzed, fell into a coma, and later died of complications. Within a year, the U.S. Gymnastics Federation, predecessor of USA Gymnastics, banned the move at levels below Olympic competition.

Deaths are rare in gymnastics. That's not the case for injuries, which Sands calls "the most pressing and serious problem faced by contemporary gymnastics." Compared with other kinds of athletics, gymnastics stands out for its singular combination of bone-jarring impacts, intense training, young age, and ever-more-demanding skills. Similar injuries afflict both sexes, but more girls participate and they start training younger. Experts fear that elite teens and preteens, by pushing their bodies to the limits, might be raising their risk of osteoarthritis and other health problems later in life. "Kids may be sustaining injuries that will be with them for a long time," says Lyle Micheli, an orthopedic surgeon at Children's Hospital in Boston.

There's no question that gymnastics is a punishing sport. Long hours of practice exact considerable wear and tear. An elite gymnast trains 25 to 40 hours a week, typically executing more than 250,000 "skills" a year. In the 1970s and '80s, elite gymnasts began competing at younger ages. Although that trend has stabilized, preteens are subjecting their bodies to sprains, fractures, and sometime even deformities of growing bones.

Gymnastics is also unusual in that it has become more demanding as judges revise the "Code of Points" used to score national and international competitions. In 1996, established routines became worth fewer

The agony of victory. Kerri Strug helped the U.S. team win gold by vaulting with an injured ankle.

High performance. Top gymnasts are scoring higher and competing harder—raising the risk of injury.

points, so gymnasts who wanted to outscore the competition had to swing higher into the air, execute more twists and somersaults, and otherwise raise their game. For a few years after that, the U.S. national team had problems with a serious kind of knee injury, a tear of the anterior cruciate ligament. Also raising the risk of injury, equipment has been modified for higher performance. For example, gymnasts can jump higher from new vaulting tables with more spring, allowing more air time for acrobatics.

Sports scientists say it is hard to nail down the health risks of competitive gymnastics for children. For starters, there is no reporting system for gymnastics injuries. In the one existing system, run by the National Collegiate Athletic Association, gymnastics usually ranks in the top three sports for injuries, behind football and hockey. But that doesn't shed light on the private clubs that train most gymnasts. Coaches of elite gymnasts often may be too busy to let scientists in for a close look, says Patrick O'Connor, an exercise scientist at the University of Georgia in Athens.

With fewer than 200 elite gymnasts in the United States, most research is done on less-skilled athletes. These studies back the impression that injuries are common. In a 3-year study of 79 female gymnasts aged 7 to 18, Dennis Caine, a sports-injury epidemiologist at Western Washington University in Bellingham, found that 60 girls suffered a total of 192 sprains, strains, and other injuries. The study also showed that the risk of injury was significantly higher for advanced gymnasts. That makes sense, because as gymnasts get serious, they put in longer hours, work harder, and attempt more difficult routines.

Caine and others have shown that the most commonly injured body parts in boys and girls include the lower back, shoulder, and ankle. Wrist pain is especially prevalent, says John DiFiori, chief of sports medicine and a team physician at the University of California, Los Angeles. In some cases, the distal end of the forearm bones can be damaged, stunting the radius relative to the ulna. "It can be career-ending for some kids, because weight-bearing is too painful," says Caine. Such nagging injuries can continue to
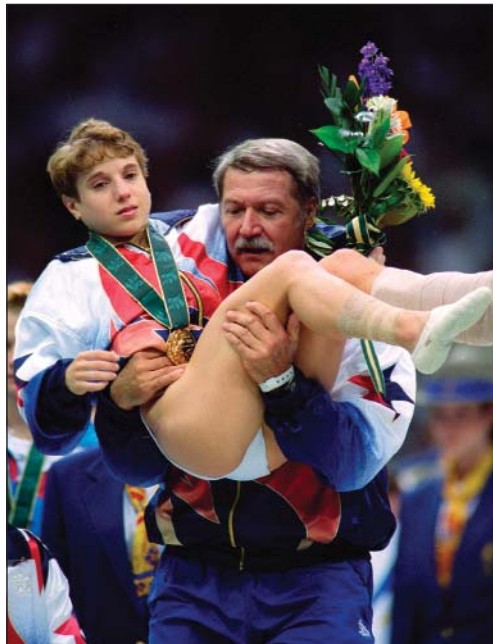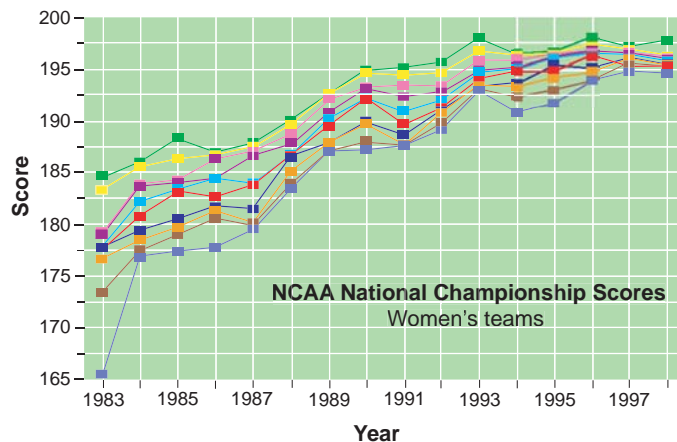
plague gymnasts in college, although he says it's unclear whether they flare up later in life.

For girls, another concern is delayed growth and sexual maturation. Elite female gymnasts tend to grow more slowly and go through puberty later than other girls. Robert Malina, a retired auxologist in Bay City, Texas, doesn't think intense training can be blamed; he says a confounding factor is that larger girls who mature earlier tend to drop out from competition. However, Caine points to several case and cohort studies that indicate at least a temporary halt in growth of some top-level gymnasts, likely due to intense training and poor nutrition, he says. When the gymnasts lightened their load, or retired, their growth rate accelerated.

Many of these issues came to a boil during the 1992 Barcelona Olympics, when European newspapers ran stories about the injuries of young female athletes and the extreme training they had undergone. That spurred a research project that many sports scientists call exceptional in its depth and quality.

The Federal Institute for Sport Science in Bonn, Germany, asked Gert-Peter Brüggemann, then director of the Institute for Athletics and Gymnastics at the German Sport University Cologne, to study the effects of high-level performance on gymnasts. His team examined various training regimens, using records from the once top-secret facilities of the former East Germany. Compared with Western gymnasts, the East Germans were put through more frequent repetitions of skill sets and more difficult maneuvers and got less rest. In total, their growing bodies had endure a longer, harder pounding compared with those of West German athletes.

Those loads had severe consequences. After reviewing archived x-rays and re-examining 42 women and 26 men who once had been elite gymnasts, Brüggemann's group found a much higher injury rate among the East German team than in 23 West Germans who had competed between 1968 and 1985. Mild deformities and abnormalities of the spine were more than twice as common in the East Germans, Brüggemann and colleague Hartmut Krahl, then of the Alfried Krupp Krankenhaus in Essen, reported in 2000 in *Belastungen und Risiken im weiblichen Kunstturnen* (*Load and Risks in Female Gymnastics*).

Seeking to prevent such injuries, the

group also carried out a 4-year prospective study of 135 young elite gymnasts on German national squads. Working with high-speed video cameras and force-measuring devices in the apparatus and landing mats, they analyzed roughly 100 exercises and measured mechanical loads exerted on the body over time. About half the spinal deformities could be explained by the amount of loading. They also discovered that the severity tended to be worse among those with weaker muscles and connective tissue. Stronger muscles absorb

Impact. The extreme forces in gymnastics, especially landings, can take a toll.

the shock of impacts and bad landings, protecting joints and the spine.

Based on those findings, Brüggemann's team devised a healthier training regimen in which girls spent less time learning fancy routines, instead logging more hours in the weight room. Far fewer gymnasts suffered pain or injury during the first 3 years of the strength-training program, they found, with ankle injuries alone falling more than 50%. One beneficiary of the program is Brüggemann's daughter Lisa, who competed in the 1999 and 2001 World Championships and will be heading to Athens on the German team. "As long as she's in this [training] system, I feel comfortable," Brüggemann says.

A similar approach has helped the U.S. national team, which has suffered fewer knee injuries since putting a greater emphasis on strength training in 2000. "We work on fitness and health rather than just skills," says team physician Lawrence Nassar of Michigan State University in East Lansing. Fitness and health are the best predictors of who will make the Olympic squad, he says.

Jill McNitt-Gray, who studies biomechanics at the University of Southern California in Los Angeles, is also working to make gymnastics safer. "I got into this because I was a gymnast and a coach," she says. "I saw too many athletes getting hurt, and I thought

there must be a better way." In her lab, McNitt-Gray takes high-speed video of gymnasts, measures forces and muscle movements, and creates computerized stick models to test the effects of modifying the moves. The gymnasts are outfitted with tiny sensors to track their motions and muscle use. This system can be used to warn trainers if the gymnasts are nearing their limits, McNitt-Gray says: "If they can't control [the motions], they're more likely to get injured."

Also on the agenda is improving the equipment. By filming the vault board and spring floor with a high-speed camera, Sands and his colleagues found that it wobbles underfoot before heaving the gymnast up. "It looks like the gymnast has landed on a waterbed," he says. That's part of what makes the equipment risky, because gymnasts must cope with boards that often compress unevenly and unpredictably. Boards have improved over the years but are still not good enough, says Sands. Similarly, the characteristics of the safety mats that gymnasts land on depend on their construction, age, and other factors. Brüggemann and his colleagues have recommended that safety mats be stiffened for better stability during landing.

Most competitive gymnasts cannot avoid an occasional injury, but little is known about the long-term damage to their health. Anecdotal evidence suggests that elite gymnasts risk developing osteoarthritis and chronic musculoskeletal pain. "In most sports, if you compete at a high level you're going to carry some baggage with you for a long time," Sands says. The few studies that have tested this hypothesis have yielded conflicting results.

Experts insist that gymnastics can be made safer without diminishing its elegance and power. "It is not as much an art as people like to think. We've got plenty of science to do and plenty of well-understood tools to find out what we need to know," says Sands. "The problem is lack of money, courage, and commitment." Perhaps the most important message is "not to cross the pain threshold," says Caine. "The old adage 'No pain, no gain' is inappropriate when it comes to kids." In the pursuit of Olympic gold, however, that message can easily get lost.

–ERIK STOKSTAD

NEWS

Engineering Peak Performance

Mechanical engineer Mont Hubbard can tell athletes how to do it faster, higher, and farther. But will they listen?

DAVIS, CALIFORNIA—Mont Hubbard handles a weighty discus the way an anthropologist might examine a 10,000-year-old skull, turning it gingerly in his fingertips. "It rolls out of the thrower's hands spinning this way," he says, turning the shiny blue "artifact," which might break your foot if it fell on you, in ultraslow motion. With his gray beard and gangly build, the 61-year-old looks nothing like an elite athlete as he rocks back in his well-worn chair in an office cluttered with baseball bats, Frisbees, and thick binders labeled "shot put," "fly casting," and "pole vault." Yet Hubbard knows how to hurl a javelin as far as it can go, guide a speeding bobsled most efficiently through a hairpin turn, and knock the longest home run.

Whereas athletes rely on countless hours of practice to master their sports, Hubbard deduces winning ways from his prodigious understanding of mechanics. "Almost all sports is mechanics," Hubbard says. "Always there are things that are moving, and almost always the motion is central. You want to know how to get something to move in a certain way, or how to get it to go farther." When it comes to analyzing and optimizing motion, Hubbard and his group at the University of California, Davis, got game. Employing basic physics and sophisticated mathematical tools, they calculate what athletes try to feel: the best way to move.

Hubbard has studied, among other things, the mechanics of the high jump and the pole vault, the aerodynamics of the discus and the shot put, the flight of punted footballs, the bend of curveballs, the flexing of fishing rods and racehorse forelimbs, the swiveling of skateboards, and the rolling of basketballs around rims. In the 1980s, he developed a system to help javelin throwers launch their spears at just the right angles. In the '90s, he built a bobsled simulator for U.S. Olympians. Currently, Hubbard and his group are analyzing ski jumping, women's gymnastics, trampoline, bungee jumping, and Frisbee flight. Last year, he kicked up controversy by arguing that a baseball batter can knock a slower-moving curveball farther than a fastball (see sidebar on p. 644).

"Hubbard is perhaps the leading figure in determining optimal strategies in sports," says William Stronge, a mechanical engineer at the University of Cambridge, U.K., who has studied the dynamics of bouncing balls. Neville de Mestre, an applied mathematician at Bond University in Gold Coast, Australia, who has done research on fluid dynamics in sports, says Hubbard is exceptionally eager to put his ideas to the test. "Many people talk about simulations," he says, "but Mont actually builds experiments to see how it works."

Hubbard's work may be of gold-medal quality to his peers, but he has had a hard time winning over coaches and athletes. "Most people in sports don't think it's useful," he admits. That may change, however, as Hubbard and other practitioners of "sports engineering" develop more powerful and realistic analyses that can better predict the keys to victory.

For the fun of it

Walking into Hubbard's Sports Biomechanics Laboratory on the Davis campus is like stepping into a strange high-tech equipment locker, minus the stench of sweaty socks. On the back wall of the windowless room
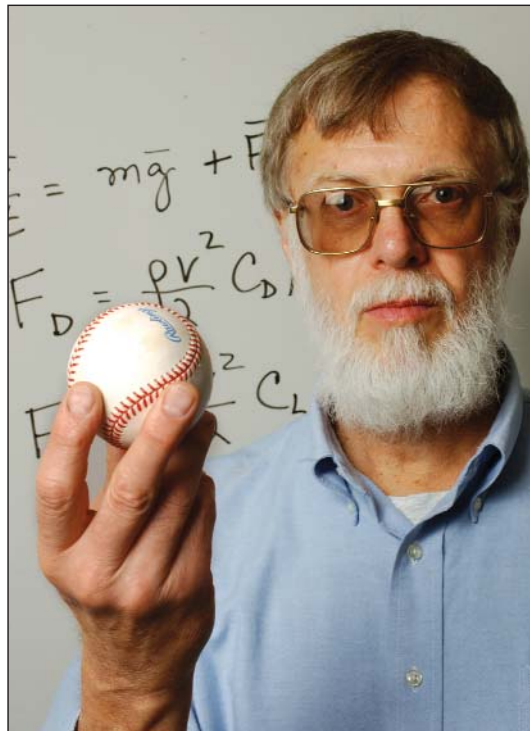
Having a ball. Engineer Mont Hubbard applies his expertise in mechanics to sports "because it's fun."

hangs a rack of javelins, their menacing points ready to perforate the air. Toward the front of the room, a blue fiberglass tub is perched on a hulking metal frame—an early version of the bobsled simulator. Nestled next to it, looking something like a homemade respirator, stands a one-of-a-kind pitching machine that can launch baseballs at 240 kilometers per hour spinning in any direction. A dent in the wall attests to the punch that such supercurveballs pack.

Hubbard has toys that any weekend warrior would envy, but he is far from a natural-born jock. "I love sports," he says, but "I was never a very successful athlete. I was always a scrawny kid." Growing up in Altavista, a small town in central Virginia, Hubbard played baseball fanatically but never blossomed into a star. He didn't really develop physically until he entered the U.S. Military Academy at West Point, New York, from which he graduated in 1964. Later in life, Hubbard played squash avidly until his second back surgery sidelined him for good.

After a stint in the Air Force, Hubbard earned his doctorate in 1975 from Stanford University in California, doing his thesis research on control systems of automobile engines. For the first several years of his career, Hubbard stuck to conventional engineering. Then, during a sabbatical at the University of Cambridge in 1981, he became fascinated with the mechanics of walking. He soon began applying his engineering skills to other types of human motion, especially sports. "It would have been an awfully boring 30 years if I'd just been doing the same old stuff," he says.

Following his heart has meant abandoning traditional funding sources. "Almost nobody pays us to do this," he says. Hubbard and his students make do primarily with four-figure grants from the U.S. Olympic Committee and various sports federations. So they look for ways to learn more by spending less. For example, in one corner of the lab a tennis ball half-covered with aluminum foil hangs from two pieces of string. Albert Jordan, an undergraduate, sets the ball spinning and swinging in front of a radar gun. He hopes that radar waves reflected by the half-silvered ball will reveal the ball's spin—a first step toward tracking the spin of a batted baseball. It's also, Jordan says, "a way to do an experiment without spending any money."

Despite the relatively spartan condi-

tions, Hubbard has no trouble attracting students. Master's degree student John Kockelman once owned a bungee-jumping company and performed stunts for television commercials; he's studying the properties of bungee cords. Doctoral student Alison Sheets competed in gymnastics from childhood through college; she's studying the dynamics of the women's uneven bars. "My brain is far better at gymnastics than my body ever was," Sheets says. "I can still do the skills, and sometimes I feel like I'm doing them better because I definitely understand them a lot better now."

Game plan

Although Hubbard has eclectic tastes in research problems, he approaches them all with a well-defined method. First, he tries to tease out the essential elements and features of an activity. For example, in analyzing the flight of a ski jumper, Hubbard and colleagues model the jumper as an assemblage of weighty links representing skis, legs, torso, and head. Using the mathematical methods of mechanical engineering, the researchers crank out differential equations that describe the forces and motions of the various parts, which they generally solve with computers. They systematically alter the variables that the athlete might control—such as the angle between a ski jumper's skis and the wind whipping past—to determine optimal values. Finally, if all goes well, Hubbard's team devel-
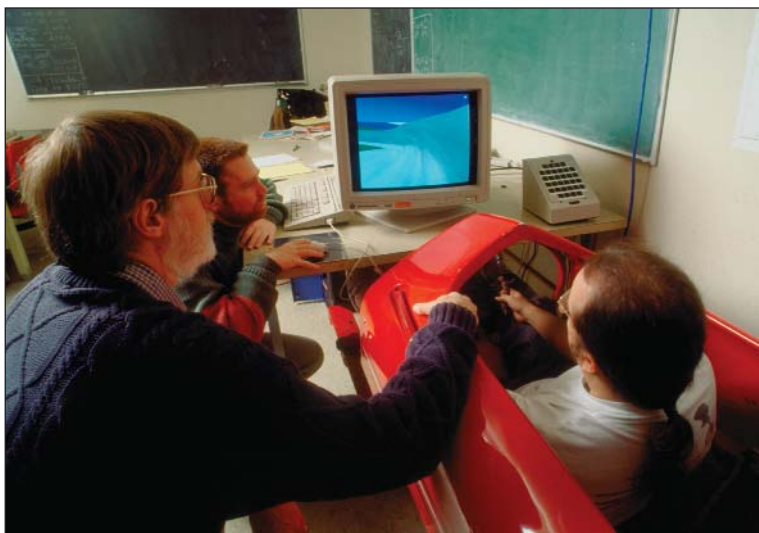
Virtually flying. Hubbard and his team tune up a bobsled simulator.

ops a methodology to help athletes reach peak performance.

Hubbard has applied this approach to myriad sports, most successfully to the throwing events in track and field. In the shot put, discus, javelin, and hammer throw, a thrower can adjust only a handful of angles and rates of rotation. Given an athlete's basic strength and speed, an engineer can calculate exactly how far the athlete will ever be able to heave the object. "That's a very beautiful concept to me," Hubbard says. "In a sense, you're allowing people to achieve the very best they can, given their physical limitations."

But such ruminations also underscore the appeal of muscle-building anabolic steroids, he says. Once an athlete has perfected a technique, the only way to throw farther is to boost strength and arm speed—for some, a pursuit that knows no bounds.

Athlete attitudes

How athletes feel about Hubbard's methods depends on whom you ask. His javelin-training system provided valuable information, says Donna Mayhew, a seven-time U.S. national champion and an Olympian in 1988 and 1992. The javelin "is such a technical sport, any feedback you can get really helps a lot," she says. "Little changes in technique are going to make big changes in distance." Hubbard's analysis showed that Mayhew tended to throw the javelin with its nose pointed several degrees too high, she says.

Bobsled driver Brian Shimer, a five-time Olympian who steered the United States to a bronze medal in 2002, says that the bobsled simulator is "great for off-season training and for working on hand-eye coordination." But the simulator cannot reproduce the feel of a bobsled ride, especially the crushing accelerations in the turns. So riding it might actually dull a driver's well-honed touch, says Shimer, who now coaches U.S. bobsled drivers. "Once the season starts," he says, "I would *not* want to get in a simulator."

And in the tradition-steeped world of baseball, scientific analysis takes a back seat to good old-fashioned coaching. Hubbard's study of hitting curveballs and fastballs makes "interesting conversation," says Gary Matthews, batting coach for the Chicago Cubs, but it won't change his coaching style. "It's more important to the scientists than it would be to me," he says. "What's important to me is that a guy has a lot of heart and can move the fastball."

Still, Hubbard and other sports engineers say they are slowly making inroads with coaches and athletes—and their numbers, although small, are growing. "There's a society of sports engineers that just didn't exist 10 years ago," says Rod Cross, a physicist at the University of Sydney, Australia, who studies tennis. This September, Hubbard and his colleagues at Davis will host the Fifth International Conference on the Engineering of Sport.

Hubbard says he doesn't worry how—or if—athletes use his ideas. He's too busy moving from project to project. "I don't want to be teaching people how to throw the javelin," Hubbard says. "I want to be thinking about new stuff." For engineers at the top of their form, that's the whole point of the game.
–ADRIAN CHO

Long Gone or Gone Wrong?

Last November, Mont Hubbard and colleagues argued in the *American Journal of Physics* that a well-hit curveball would sail farther than a perfectly struck fastball—even though the curveball moves slower and packs less energy, both before and after it's walloped. That's because a batted ball travels farther if it has more backspin to give it aerodynamic lift. The top-spinning curveball approaches the batter already turning in a direction that increases the backspin of the batted ball. On the other hand, the back-spinning fastball comes at the batter spinning the wrong way, which decreases the backspin of the outgoing orb. As a result, an optimally struck curveball will travel around 455 feet (138 meters), about 12 feet (3.5 meters) farther than a well-hit fastball.

Not so, contends Robert Adair, a physicist at Yale University and author of *The Physics of Baseball*. Although he can't say precisely what's wrong with Hubbard's calculations, he claims that balls moving at the speeds Hubbard quotes just don't go that far, so Hubbard and colleagues must have overestimated the lifting effect of spin. For his part, Gary Matthews, batting coach for the Chicago Cubs, says Hubbard's team may be right—especially if the curveball "hangs" high in the strike zone. "A hanging curveball will go a long, long way," he says. Although Matthews may not be an expert in aerodynamics, he did belt 234 homers in 16 seasons in the major leagues. **–A.C.**

APOBEC-Mediated Editing of Viral RNA

Kate N. Bishop, Rebecca K. Holmes, Ann M. Sheehy, Michael H. Malim*

Viral sequence variation frequently plays an essential role in viral replication and pathogenesis. For instance, sequence changes can underlie alterations in viral phenotype, escape from host immune responses, and the acquisition of drug resistance. Usually, such changes are due to the fixation of copying errors made during genome replication. Sequence variation may also arise through the post-synthesis editing of RNA or DNA: Measles, hepatitis delta, and polyoma viral RNAs, for example, are presumed substrates for adenine deamination (1). Editing of human immunodeficiency virus (HIV) DNA by the human cytidine deaminase APOBEC3G (hA3G) has also received much recent scrutiny. HIV particles carrying this enzyme yield minus-strand reverse transcripts in which multiple cytosine (C) residues are deaminated to uracil (U). The resulting hypermutation [guanine (G) to adenine (A) on plus strands] and cDNA instability combine to inhibit viral infectivity. Normally, the viral protein Vif protects HIV from hA3G by inducing its degradation and exclusion from virions (2).

Beyond hA3G, mammalian genomes harbor a variety of additional RNA/DNA cytidine deaminases (3). The founder member of this APOBEC family is APOBEC1 (human A1, or hA1), the catalytic subunit of a complex that deaminates C^{6666} in apolipoprotein B mRNA in human gastrointestinal tissue to yield a premature stop codon. To determine whether other APOBEC proteins can mutate HIV and/or regulate infectivity, wild-type or vif-deficient (△vif) viruses were produced in the presence of various family members, and viral infectivities were measured (Fig. 1A) (4, 5). As expected, hA3G profoundly inhibited the △vif virus but had a mild effect on its wild-type counterpart. Surprisingly, whereas hA1 had negligible effects, rat A1 (rA1) was a potent suppressor of both viruses. Although the basis for the dichotomy between rA1 and hA1 (which share ~70% amino acid sequence identity) is not understood, we speculate that either hA1 is intrinsically incapable of influencing HIV infection or that 293T cells lack cellular cofactors critical for this activity (6).

Because hA3G induces excessive C-to-U deamination of minus-strand HIV cDNA, we asked whether this extended to rA1. Viral cDNAs were recovered from infected cells, and a nef/3′–long-terminal-repeat (LTR) fragment was then amplified by polymerase chain reaction (PCR), cloned, and individual sequences determined (Fig. 1B). The mutations induced by hA3G were ~95% G-to-A transitions (5). However, this was only the predominant mutation (~70%) induced by rA1, with many C-to-T mutations also being evident (~22%). These particular mutations could have arisen through deamination of unpaired plus-stranded cDNA or virion RNA.

To address the latter possibility, cell-free virions were purified and a nef/3′-LTR region of the genomic RNA was converted to cDNA in vitro, amplified by PCR, cloned, and sequenced (Fig. 1C). There was a clear accumulation of C-to-T (U in the template RNA) changes in the RNAs of HIV produced in the presence of rA1 (17 events in 23,982 bases, a frequency close to that seen with DNA sequencing). Thus, HIV RNA as well as DNA can be a substrate for APOBEC-mediated deamination, making it plausible that C-to-U editing of either could contribute to the sequence diversity that marks natural HIV infection (7). In both cases, this will require expression of APOBEC proteins and their presumed cofactors in the natural targets of HIV infection, namely T lymphocytes and macrophages.

These observations allow speculation (i) that APOBEC-mediated C-to-U editing may contribute to the sequence variation of viruses that replicate entirely through RNA, of which there are many; and (ii) that additional cellular RNA substrates might exist for the APOBEC enzymes. Both possibilities will require further rigorous experimental investigation. It has also been demonstrated that hA3G inhibits the association of hepatitis B virus RNA with viral cores—in this case, in the absence of viral RNA/DNA deamination (8). Thus, APOBECs have the potential to affect their substrates through at least three mechanisms: DNA editing, RNA editing, and a nonediting pathway. The possibilities for APOBEC involvement in the regulation and dysregulation of cell function (6, 7), as well as in innate resistance to invading nucleic acids, may therefore be considerable.

Fig. 1. (A) Infectivity of HIV produced in the presence of hA3G, hA1, or rA1; error bars show standard deviation. WT, wild type. (B) Plus-strand cDNA mutation spectra for the indicated virus/APOBEC pairs; n = the total number of sequenced base pairs. (C) rA1 deaminates viral RNA. Clones (n = 21) were analyzed for each virus, with all changes marked onto a single horizontal line representing the entire amplicon; C-to-T changes are indicated by asterisks.

References and Notes

1. B. L. Bass, *Annu. Rev. Biochem.* **71**, 817 (2002).
2. X. Yu *et al.*, *Science* **302**, 1056 (2003).
3. A. Jarmuz *et al.*, *Genomics* **79**, 285 (2002).
4. Materials and methods are available as supporting material on *Science* Online.
5. K. N. Bishop *et al.*, *Carr. Biol.*, in press; published online 24 June 2004 (10.1016/50960982204004683).
6. S. Anant, N. O. Davidson, *Trends Mol. Med.* **9**, 147 (2003).
7. R. C. Beale *et al.*, *J. Mol. Biol.* **337**, 585 (2004).
8. P. Turelli, B. Mangeat, S. Jost, S. Vianin, D. Trono, *Science* **303**, 1829 (2004).
9. Supported by the U.K. Medical Research Council, the Biotechnology and Biological Sciences Research Council, the Royal Society (A.M.S.), and the Elizabeth Glaser Pediatric AIDS Foundation (M.H.M.).

Department of Infectious Diseases, Guy's, King's and St. Thomas' School of Medicine, King's College London, London, SE1 9RT, UK.

*To whom correspondence should be addressed. E-mail: michael.malim@kcl.ac.uk

Engineered Interface of Magnetic Oxides

Hiroyuki Yamada,[1,2] Yoshihiro Ogawa,[3] Yuji Ishii,[1,2] Hiroshi Sato,[1,2] Masashi Kawasaki,[1,4] Hiroshi Akoh,[1,2] Yoshinori Tokura[1,3,5]

Interface-selective probing of magnetism is a key issue for the design and realization of spin-electronic junction devices. Here, magnetization-induced second-harmonic generation was used to probe the local magnetic properties at the interface of the perovskite ferromagnet $La_{0.6}Sr_{0.4}MnO_3$ with nonmagnetic insulating layers, as used in spin-tunnel junctions. We show that by grading the doping profile on an atomic scale at the interface, robust ferromagnetism can be realized around room temperature. The results should lead to improvements in the performance of spin-tunnel junctions.

In the development of semiconductor technology, a major role was played by the elucidation of electronic states at heterointerfaces such as positive-negative (p-n) junctions and metal/semiconductor interfaces. In the p-n junction, for example, a depletion layer and built-in electric field are formed at the interface so as to unify the Fermi level all over the junction. Analogously, transition-metal-oxide compounds, as emergent electronics materials with correlated electrons, often exhibit nontrivial and potentially useful electronic properties at heterointerfaces. In most cases, however, such interface properties cannot be explained in

terms of conventional band pictures (1). Moreover, the charge state in the strongly correlated electron oxides is closely coupled with the spin and orbital degrees of freedom. Therefore, magnetic oxide interfaces provide a challenging arena to explore new multifunctional materials (2).

Interface magnetism is also important for the development of devices with strongly correlated electron oxides. For example, the perfectly spin-polarized ferromagnet $La_{1-x}Sr_xMnO_3$ (LSMO) could be one of the best candidates for magnetic random-access memory if we could fully use its potential in spin-tunnel junctions. However, in the actual junctions—for example, LSMO/$SrTiO_3$ (STO)/LSMO—the tunnel magnetoresistance (TMR) has been much smaller than that expected from the half-metallic nature of LSMO, and the response has been diminished at temperatures far below T_c (3, 4). The possible origin of the decreased MR is severe deterioration of the ferromagnetism locally occurring near the LSMO/STO interfacial region (5–7). Therefore, selective and quantitative evaluation of interfa-

cial spin state may provide a key to realizing high-performance magnetic devices.

Nonlinear magneto-optical effects can probe the interface magnetism (8, 9), as has recently been demonstrated for oxide "tricolor" superlattices, where ferromagnetic LSMO layers and insulating STO and $LaAlO_3$ (LAO) layers were accumulated one by one in an ABCABC. . . structure. The STO/LSMO/LAO superlattice exhibits very large magnetization-induced second-harmonic generation (MSHG) and a resultant nonlinear magneto-optical Kerr effect (NOMOKE) (10, 11). This "tricolor" spin superlattice is regarded as an artificially constructed noncentrosymmetric ferromagnet, where symmetries of space inversion and time reversal are both broken simultaneously. The generic feature of such a noncentrosymmetric ferromagnet should be extended to various heterointerfaces of strongly correlated electron oxides.

We show that such an interface magnetism can indeed be probed by MSHG from single heterojunctions between LSMO and STO or LAO, and we have found a robust ferromagnetic interface of LSMO. It can be obtained by modulating the interface doping profile on an atomic scale, which may open up a new path to LSMO-based spin-injection devices such as TMR junctions operating at room temperature.

The schematic side view of the heterostructures used in the present study is depicted in Fig. 1A. For detecting the interface magnetization at a single magnetic interface, we prepared bilayer films composed of an insulating cap layer (STO or LAO) with a thickness of 5 unit cells (u.c.) (i.e., 2 nm) and a magnetic LSMO bottom-layer grown on STO(001) substrates (Fig. 1A, left). The LSMO layer is as thick as 300 u.c (120 nm). Because the optical absorption coefficients in LSMO for incident beam (λ = 800 nm) and SH light (λ = 400 nm) are about 1×10^5 cm^{-1}, the contribution from the bottom film/
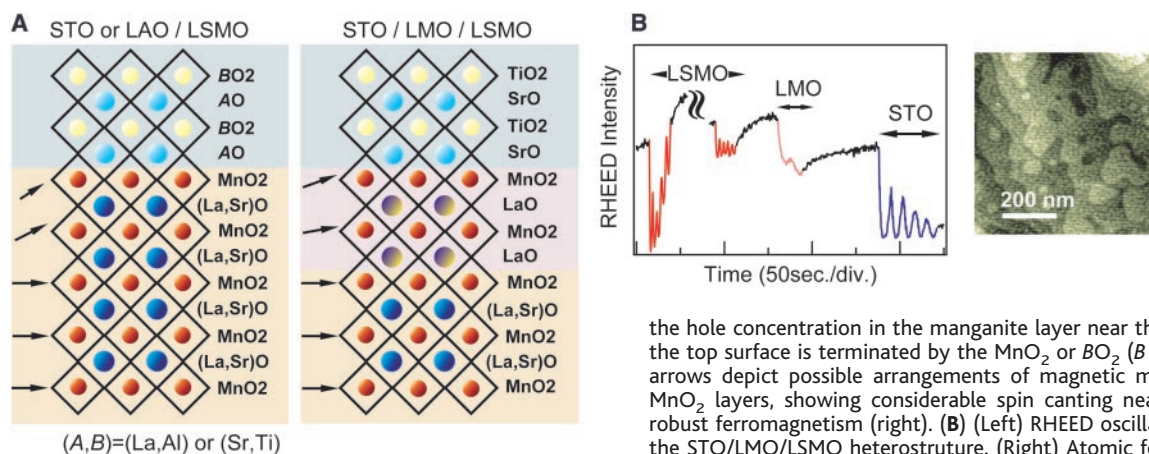
Fig. 1. (A) Atomic stacking sequences for perovskite-type magnetic oxide interfaces. (Left) LSMO film capped with a 5-u.c. (2-nm) layer of $LaAlO_3$ (LAO) or $SrTiO_3$ (STO). (Right) A 2-u.c. layer of undoped LMO is inserted between STO and LSMO layers to grade the hole concentration in the manganite layer near the interface. In these films, the top surface is terminated by the MnO_2 or BO_2 (B = Ti or Al) layer (18). The arrows depict possible arrangements of magnetic moments at the respective MnO_2 layers, showing considerable spin canting near the interface (left) and robust ferromagnetism (right). **(B)** (Left) RHEED oscillation during the growth of the STO/LMO/LSMO heterostructure. (Right) Atomic force microscopy image for the top surface of STO/LMO/LSMO. Atomically flat terraces with the single-u.c. steps (0.4 nm high) are clearly seen.

substrate interface is negligibly small (<1%). All the films were fabricated by pulsed laser deposition (PLD) in a layer-by-layer growth mode while observations were made of the intensity oscillation of the specular spot in reflection high-energy electron diffraction (RHEED) (Fig. 1B). The resultant film surfaces are atomically flat with 200-nm-wide terraces and 0.4-nm-high steps. We also note that there should be no detectable interface roughness or disorder, judging from the transition electron microscopy image and x-ray diffraction for the "tricolor" superlattice grown in a similar way (12). Therefore, we may conclude that exact control of interface atomic stacking can be realized. Four-circle x-ray diffraction measurements indicate that the in-plane lattice constants are identical with that of the substrate (STO), ensuring coherent epitaxy throughout the bilayer film. Magnetization of the bilayer film (M_{film}) was measured with a superconducting quantum interference device (SQUID) magnetometer, indicating in-plane magnetic anisotropy, well-defined T_c of 340 K, and saturated magnetization of 3.6 μ_B / Mn for all the samples.

Figure 2A illustrates the experimental configuration for the MSHG measurements (13). The external or built-in fields induce magnetization (M), electric polarization (P), and toroidal moment ($T \equiv P \times M$), which gives MSHG (Fig. 2A, bottom). The SH light intensity (I), measured in a magnetic field (H) of ± 0.05 T, was plotted against the analyzer angle (θ), as shown in Fig. 2C. MSHG cor-

responds to the $s(y)$-polarized component ($\theta = 90°$) in the SH light. A phase shift in the I-θ curves for ± H gives twice the NOMOKE angle, from which we can deduce the intensity of MSHG. We define the square root of the MSH intensity as the interface magnetization (M_{int}), which is plotted against T in Fig. 2B. In the upper panel, the M_{int}-T plot is normalized with M_{int}(50 K) and compared with the M_{film}-T curve.

The MSHG data shown in Fig. 2C indicate a large variation of magnetic properties between STO/LSMO and LAO/LSMO interfaces. The STO/LSMO interface hardly shows MSHG or NOMOKE even at 50 K, which directly demonstrates the existence of a dead layer at the interface and explains the inferior TMR response found in LSMO/STO/LSMO junctions (5). At the LAO/LSMO interface, by contrast, large MSHG is observed, and the ferromagnetism survives even at 250 K, although the onset T of the MSHG signal as the interface T_c is not so high or well defined as the T_c of the bulky thick films. The present observation is consistent with the finding of the studies on transport properties of the related superlattices that the LAO and STO form robust and susceptible interfaces with LSMO, respectively (10, 11).

Figure 3 exhibits the magnetic field dependence of M_{int} as deduced by the square root of MSHG intensity, measured at T = 50 K. This "interface M-H curve" enables us to decompose M_{int} into the spontaneous magnetization and field-induced part. To clearly distinguish the

two components, M_{int} is fitted with the spontaneous magnetization plus the H-linear function and normalized with M_{int}(7 T) (Fig. 3, top). The magnetic field of 7 T increases the spontaneous magnetization by 200% and 70% for STO/LSMO and LAO/LSMO interfaces, respectively. This means that the antiferromagnetic spin canting occurs at the STO/LSMO interface, while the ferromagnetic spin arrangement is much less canted by LAO.

In bulk crystals, LSMO (x = 0.4) has the highest T_c of 370 K among the La$_{1-x}$Sr$_x$MnO$_3$ series (14), but the LSMO (x = 0.4) film grown on STO is on the verge of an A-type (layered) antiferromagnet with reduced T_c(340 K) as a result of the tensile strain (15). This instability toward the antiferromagnetic phase explains why the ferromagnetic spin ordering in LSMO (x = 0.4) is so easily reduced at the interface. Moreover, STO produces a dead layer in the adjacent LSMO more aggressively than does LAO, because the STO layer works as a hole-donating layer. It has been considered that the valence-mismatched interface composed of the stacking sequence -TiO$_2$-SrO-MnO$_2$-La$_{0.6}$Sr$_{0.4}$O- induces the charge transfer and that the overdoped LSMO is dominated by the antiferromagnetic spin canting (5–7, 10, 16). Therefore, the underdoped LSMO (preferably x < 0.3) layer may enable us not only to stabilize the ferromagnetism but also to compensate the charge transfer from STO at the interface. However, both high T_c and conductivity are lost by decreasing x, where we confront a dilemma.
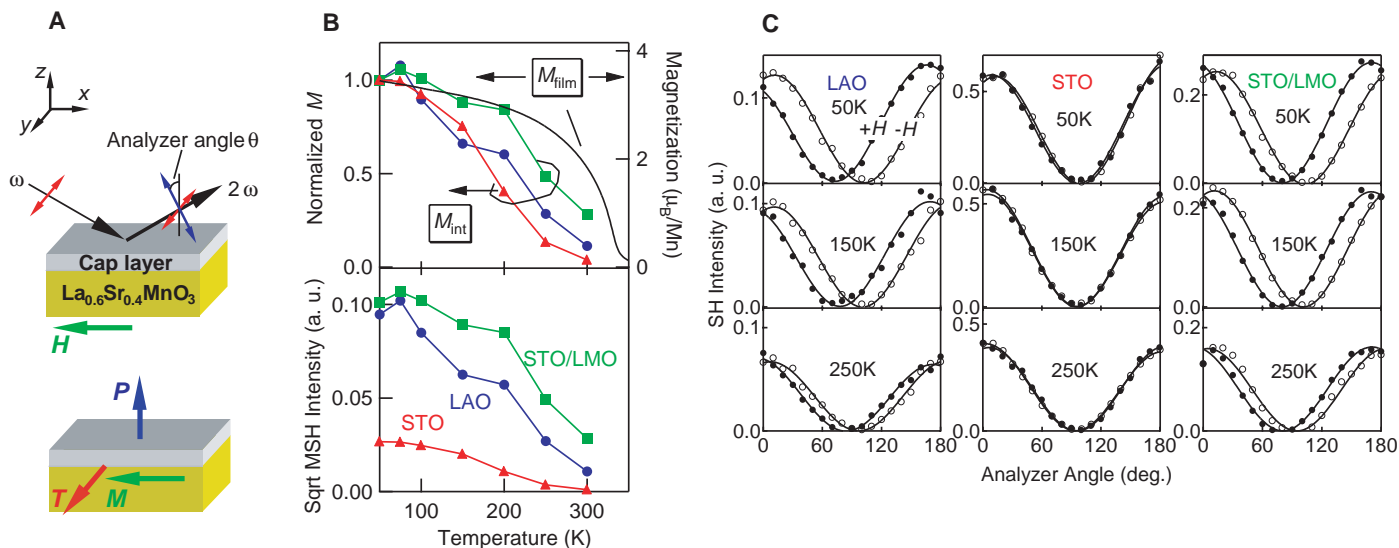

Fig. 2. (A) (Top) Schematic of experimental configuration for the measurement of MSHG. Polarization of incident light and applied magnetic field (H) are along y and x, respectively. The reflected SH light undergoes Kerr rotation. (Bottom) Configuration of various electric and magnetic moments. P (//z) stands for electric polarization originating from non-centrosymmetric atomic stacking along the z direction. M (//x) shows magnetization of LSMO along the film plane induced by a magnetic field H (//x). T (//y) indicates a toroidal moment, which is generated as $T = P \times M$. **(B)** (Bottom) Temperature (T) dependence of magnetization at the interface (M_{int}), defined as the square root of MSHG intensity (s-in s-out

SH light). The applied magnetic field is 0.05 T. Red, blue, and green symbols indicate the data for STO/LSMO, LAO/LSMO, and STO/LMO/LSMO interfaces, respectively. (Top) $M_{int}(T)$ normalized with M_{int}(50 K), and T dependence of magnetization of the films (M_{film}) measured at 0.05 T with a SQUID magnetometer (a black line). **(C)** SH intensities plotted against analyzer angle (θ) at 50 K, 150 K, and 250 K for +H and –H. θ denotes 0° and 90° for p- and s-polarized SH light, respectively. The θ dependence can be fitted to the formula $|P_z\cos\theta \pm P_y\sin\theta|^2$. Phase shift induced by the H reversal corresponds to twice the NOMOKE angle, which gives the magnitude of M_{int}.
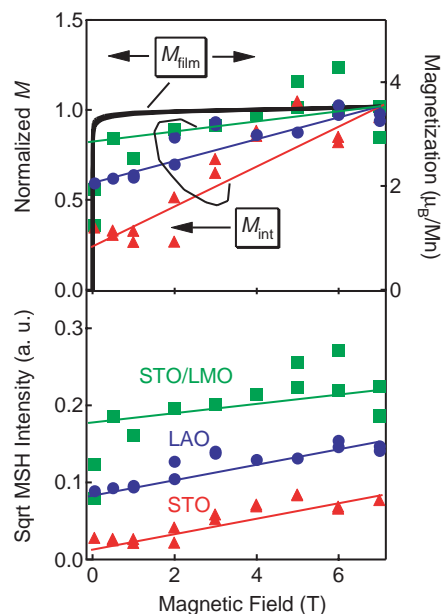
Fig. 3. Magnetic field (*H*) dependence of the square root of MSHG intensity, representing the interface magnetism (M_{int}), measured at 50 K. M_{int} (**bottom**) and normalized M_{int} [$M_{int}(H)/M_{int}(7\ T)$] (**top**). STO/LSMO, LAO/LSMO, STO/LMO/LSMO interfaces are displayed with red, blue, and green symbols, respectively. The solid lines are the fit, assuming spontaneous magnetization (intersection to the vertical axis) and *H* linear increase in M_{int} from the canted states by an external field. A black line in the top panel shows the *H* dependence of M_{film} at 5 K.

To enhance the interface magnetization without sacrificing the ferromagnetic and metallic characters in the film, we propose the compositionally graded LSMO interfaces, where the bulk LSMO electrode has the optimal doping level of *x* = 0.4, and *x* decreases gradually to 0 toward the insulating layer. This idea bears some analogy to the procedure to improve the superconductivity grain boundary junction (*17*), in which partial Ca substitution on Y sites in $YBa_2Cu_3O_{7-x}$ enhances the critical supercurrent density (J_c), but at the same time suppresses T_c. In our study, the doping profile was controlled on an atomic scale. The 2-u.c. layer of $LaMnO_3$ (LMO) was inserted as a locally underdoped layer between STO (5 u.c.) and LSMO (*x* = 0.4; 300 u.c.) layers (Fig. 1A, right). We anticipated that the LMO layer, which is originally an antiferromagnet with T_N = 140 K, should be hole-donated by LSMO(0.4) and STO layers and hence can compensate for the interface effects. This compositionally modulated interface shows large MSHG and NOMOKE, which exceed by far the STO/LSMO direct interface and are comparable to those of the LAO/LSMO interface (Fig. 2B). In accord with this observation, the spin-tunnel junction equipped with such an atomically engineered interface, [LSMO(0.4)/LMO (2 u.c.)]/STO (2 nm)/[LMO (2 u.c.)/LSMO(0.4)], shows an improved (comparable)

device performance as compared with the direct-interface junction, LSMO/STO/LSMO (LSMO/LAO/LSMO) (*13*). As shown in Fig. 3, M_{int} for this STO/LMO/LSMO interface is almost *H*-independent above 0.1 T, thus confirming that the nearly bulk-like ferromagnetism is maintained even in the vicinity of STO. The normalized $M_{int}(T)$ value of STO/LMO/LSMO reaches 0.5 at 250 K, which is close to a value (0.7) for normalized magnetization of bulk film at the same temperature and is much larger than the value (<0.2) for the direct interface of STO/LSMO (*x* = 0.4).

References and Notes
1. A. Ohtomo, D. A. Muller, J. L. Grazul, H. Y. Hwang, *Nature* **419**, 378 (2002).
2. K. Ueda, H. Tabata, T. Kawai, *Science* **280**, 1064 (1998).
3. J. Z. Sun *et al.*, *Appl. Phys. Lett.* **69**, 3266 (1996).
4. M. Bowen *et al.*, *Appl. Phys. Lett.* **82**, 233 (2003).
5. Y. Ogimoto *et al.*, *Jpn. J. Appl. Phys.* **42**, L369 (2003).
6. M. Izumi *et al.*, *Phys. Rev.* **B64**, 064429 (2001).
7. M. Izumi, Y. Ogimoto, T. Manako, M. Kawasaki, Y. Tokura, *J. Phys. Soc. Jpn.* **71**, 2621 (2002).
8. J. Reif, J. C. Zink, C.-M. Schneider, J. Kirschner, *Phys. Rev. Lett.* **67**, 2878 (1991).
9. G. Spierings *et al.*, *J. Magn. Magn. Mater.* **121**, 109 (1993).
10. H. Yamada, M. Kawasaki, Y. Ogawa, Y. Tokura, *Appl. Phys. Lett.* **81**, 4793 (2002).
11. Y. Ogawa *et al.*, *Phys. Rev. Lett.* **90**, 217403 (2003).
12. K. Kimoto *et al.*, *Appl. Phys. Lett.* **84**, 5374 (2004).
13. Materials and methods are available on *Science* Online.
14. A. Urushibara *et al.*, *Phys. Rev.* **B51**, 14103 (1995).
15. Y. Konishi *et al.*, *J. Phys. Soc. Jpn.* **68**, 3790 (1999).
16. F. Pailloux *et al.*, *Phys. Rev.* **B66**, 014417 (2002).
17. G. Hammerl *et al.*, *Nature* **407**, 162 (2000).
18. M. Izumi *et al.*, *Appl. Phys. Lett.* **73**, 2497 (1998).
19. We thank T. Arima, J. Matsuno, and A. Sawa for enlightening discussions.

Real-Space Observation of Molecular Motion Induced by Femtosecond Laser Pulses

Ludwig Bartels,[1] Feng Wang,[2] Dietmar Möller,[2] Ernst Knoesel,[3] Tony F. Heinz[2]*

Femtosecond laser irradiation is used to excite adsorbed CO molecules on a Cu(110) surface; the ensuing motion of individual molecules across the surface is characterized on a site-to-site basis by in situ scanning tunneling microscopy. Adsorbate motion both along and perpendicular to the rows of the Cu(110) surface occurs readily, in marked contrast to the behavior seen for equilibrium diffusion processes. The experimental findings for the probability and direction of the molecular motion can be understood as a manifestation of strong coupling between the adsorbates' lateral degrees of freedom and the substrate electronic excitation produced by the femtosecond laser radiation.

Scanning tunneling microscopy (STM), through its direct imaging on the atomic length scale, has dramatically advanced our ability to probe the geometric and electronic structure of surfaces. The introduction of variable temperature STM has further enabled researchers to follow atomic motion on surfaces in real time. Although this capability permits the examination of processes on the millisecond time scale, the inherent characteristics of STM imply that new approaches are required to reach the intrinsic time scale of atomic motion, which is in the pico- to femtosecond range.

Several researchers have explored gating and correlation techniques to improve the time resolution of STM measurements (*1–3*), but their focus has been on electron dynamics rather than on the nuclear motion that corresponds to chemical processes. In a separate line of research, the utility of femtosecond laser radiation to induce surface chemical reactions and to clarify rates of energy transfer and nuclear motion has been demonstrated (*4–14*). Although they access the ultrafast dynamics of surface reactions, these methods do not provide the direct real-space imaging of STM.

We describe here the successful combination of the atomic-scale spatial resolution provided by STM with ultrafast surface dynamics driven by femtosecond laser excitation. In this fashion, we are able to determine the initial and final configurations of individual molecules undergoing nonequilibrium surface diffusion induced by electronic exci-

[1]Pierce Hall, University of California, Riverside, CA 92521, USA. [2]Department of Physics and Department of Electrical Engineering, Columbia University, New York, NY 10027, USA. [3]Department of Physics and Astronomy, Rowan University, Glassboro, NJ 08028, USA.

*To whom correspondence should be addressed. E-mail: tony.heinz@columbia.edu

tation. Given the critical role that surface diffusion plays in phenomena as varied as crystal growth and catalytic activity of surfaces, many techniques have been applied to probe the diffusion of adsorbed atoms and molecules under equilibrium conditions (15, 16). By exciting the molecules electronically rather than thermally, we can now examine surface diffusion processes occurring under nonequilibrium conditions on the ultrafast time scale. The measurements, carried out on the CO/Cu(110) system, reveal adsorbate motion with different propensities for displacement along and perpendicular to the atomic rows of the surface compared with those previously observed under conditions of thermal equilibrium. The experimental results demonstrate the importance of the electronic adsorbate-substrate coupling for surface migration, as well as the existence of distinctive new surface phenomena on the ultrafast time scale.

We chose the CO/Cu(110) system for study because its equilibrium and dynamical properties have been well characterized. CO adsorbs in a molecular state at on-top sites of the Cu(110) surface and is stable against thermal desorption up to a temperature of ~190 K (17). With the use of STM imaging, Briner et al. (18) found that thermally induced diffusion starts at temperatures around 40 K and proceeds exclusively along the close-packed <1$\bar{1}$0> rows of the substrate. Researchers have also recently investigated STM-induced lateral motion of adsorbed CO (19–23). These measurements in the spatial domain have been complemented by experimental investigations of femtosecond laser-induced processes, including research on desorption (5, 7) and adsorbate-substrate vibrational coupling (24–27), as well as several related theoretical studies (28–31). These experimental and theoretical investigations have revealed the importance of nonadiabatic coupling of the adsorbate to the high electronic temperature of the substrate produced by femtosecond laser excitation.

During the experiment (32), the Cu(110) sample was held at a temperature of 22 K, well below the threshold for thermally driven surface diffusion of CO. The CO coverage of ~0.01 monolayer permitted isolated molecules to be studied. STM images were taken at tunneling currents of ~100 pA with a bias of a few hundred millivolts. Under these conditions, no STM-induced changes of the CO-covered surface were observed. Laser radiation was provided by a regeneratively amplified mode-locked Ti:sapphire laser. We used the second harmonic of this source (405 nm) to irradiate the sample with 200-fs laser pulses at a repetition rate of ~1 kHz.

We recorded the positions of CO molecules on the Cu(110) surface by STM before and after irradiation by the laser and thereby traced the motion of individual adsorbed molecules. Although conceptually simple, this procedure requires careful consideration from the experimental standpoint. The principal complication is that the surface cannot be irradiated with the STM tip in its tunneling position because of thermal expansion.

The experimental approach adopted to circumvent this problem is sketched in Fig. 1. After acquisition of an initial STM image, the tip is quickly withdrawn from the substrate by ~1 μm with the vertical piezoelectric drive. With the tip retracted, the surface is exposed to the fs-pulse laser radiation. The STM tip is then allowed to reapproach the surface. To use the STM to follow atomic-scale motion, we must be able to image precisely the same region of the surface. Control experiments, as indicated in Fig. 2, A and B, confirm this capability and verify that the molecules retain exactly the same positions. In contrast, Fig. 2, B and C, show STM images of adsorbed CO molecules before and after exposure to fs-pulse laser radiation. Because the probability of laser-induced events is low for a single pulse, we typically applied 1000 or more pulses between successive images. In our measurements, a total of ~3000 isolated molecules were

analyzed, half of them subject to laser irradiation and half in control experiments.

Under irradiation by the femtosecond laser pulses, we observed surface diffusion as summarized in Fig. 3. In contrast to the thermal behavior, adsorbate migration occurs across the close-packed rows of the substrate (in the <001> direction) as well as along them (in the <1$\bar{1}$0> direction). Also in marked distinction to conventional thermal behavior, desorption occurs at a rate similar to surface diffusion.

An important experimental issue concerns the determination of the absorbed laser fluence, which has been shown to be a critical experimental parameter (4–13). The geometry of the present experiment makes a direct determination of the absorbed fluence difficult: Even with the STM tip retracted, we expect that its presence will alter the strength of the laser field at the surface. We can, however, apply an internal calibration scheme on the basis of the presence of laser-induced CO desorption. To convert the observed desorption rates into an absorbed fluence, we apply the phenomenological model developed by Struck et al. (7) for Cu(001), with an adjustment to reflect the empirical Arrhenius rate law for CO desorption from Cu(110). Given the sharp dependence of the desorption rate on the absorbed laser fluence, we believe that this procedure yields a relatively precise local value of this parameter of 30 ± 3 J/m^2.

We now discuss the mechanism for femtosecond laser-induced diffusion and the origin of the anomalous branching ratio for motion in the <1$\bar{1}$0> and <001> directions. The absorbed laser radiation produces electron-hole pairs that thermalize rapidly to a Fermi-Dirac distribution. However, this electronic temperature can far exceed that of the phonons (25). To model this transient substrate excitation, we have solved coupled diffusion equations for the substrate electronic and lattice temperatures at the surface (Fig. 4). A transient electronic temperature approaching

Fig. 1. Overview of experimental procedure. (**A**) Initially, a molecularly resolved STM image of the surface is acquired. (**B**) The tip is retracted to the limit of the feedback loop, the vertical piezo is switched to a high-voltage supply, and the tip is retracted more than 1 μm from the sample. Laser radiation is then applied. (**C**) The tip is brought back toward the surface, and the feedback loop becomes operational. After creep compensation, the surface is reapproached and stable tunneling is achieved. Finally, we restart the STM imaging. In the following control cycle, the entire procedure is repeated without laser irradiation.
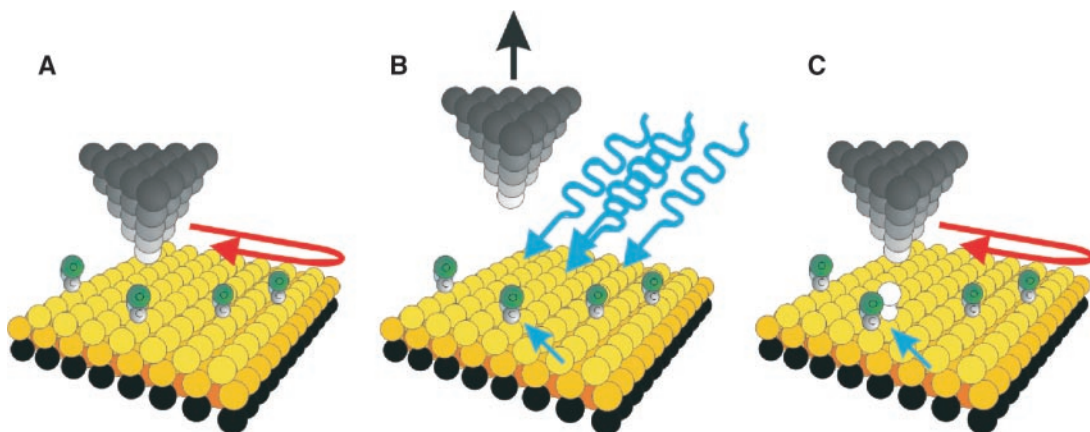
3000 K is achieved for a time of ~1 ps; heating of the lattice also occurs, but the lattice only reaches a temperature of 140 K in a trans-ient of a few tens of picoseconds in duration.

With this description of the substrate excitation, we first consider whether the ob-served diffusion arises from a transient thermal process associated with the (relatively) slow equilibrium heating in the metal. With the use of data from Briner et al. (18), we find that the expected number of hops along the atomic rows falls short by more than an order of magnitude. Also, CO hopping across the rows is negligible under conventional thermal conditions and should, contrary to our experimental findings, still be relatively minor for transient thermal excitation. Further, a purely thermal mechanism could not explain the concurrent CO desorption process observed under laser excitation.

These considerations strongly suggest that an electronically driven process is responsible for the observed surface diffusion. Although theoretical attention has focused largely on the role of electronic temperature in driving desorption processes, the notion of electronic coupling to coordinates relevant for lateral adsorbate motion is also inherently reasonable, because strong adsorbate-substrate charge transfer is expected as the adsorbate moves from one surface binding site to the next. Indeed, in the calculations of Kindt et al. (30), it was explicitly noted that electronic coupling to modes other than the molecule-surface vibration for CO/Cu(100) should play an important role.

Here, we introduce a simple phenomenological model of such an electronically driven diffusion process. With the use of only parameters from the literature, we can roughly reproduce the experimental rate for diffusion along the atomic rows of the Cu(110) surface and account for the anomalous behavior for diffusion across these rows. Our starting point is the experimental Arrhenius rate law reported by Briner et al. (18) for CO diffusion along the $\langle 1\bar{1}0 \rangle$ direction. We now, however, consider energy transfer separately from the electronic and lattice degrees of freedom of the substrate, following a model previously introduced to examine coupling between the substrate and adsorbate vibrations:

$$dU_{ad}/dt = (U_{el} - U_{ad})/\tau_{el} + (U_{ph} - U_{ad})/\tau_{ph}$$

(1)

where $U_x = h\nu/\exp(h\nu/kT_x - 1)$ denotes the energy content of a harmonic oscillator corresponding to the frustrated translational mode ($\nu = 1.1 \times 10^{12}$ s^{-1}) at the specified temperature T_x (where ad, el, and ph indicate the adsorbate, electronic, and phonon degrees of freedom, respectively) and h and k are the usual Planck and Boltzmann constants. The coupling parameters ($\tau_{el} = 5.1$ ps and $\tau_{ph} = 4.2$ ps) for frustrated translational motion, which is assumed to be the reaction coordinate for surface diffusion, were obtained experimentally by Germer et al. for the CO/Cu(001) system (24, 25). Although we have included for

Fig. 2. Portion of STM images (1.00 V, 200 pA, and 50 Å by 25 Å) of CO molecules (dark regions indicated by circles) on the Cu(110) surface taken at high scan speed. The rows of surface atoms along the $\langle 1\bar{1}0 \rangle$ direction are highlighted on the basis of separate slow scans. The tip was removed and reapproached without optical excitation between images (**A**) and (**B**). Although a slight drift is seen, precisely the same region of the surface has been found. Between images (B) and (**C**), fs-pulse laser excitation was applied. (C) has been numerically corrected for drift. One adsorbed CO molecule (with an initial position indicated by the dashed yellow circle) moved to a new position (red circle) one site along the $\langle 1\bar{1}0 \rangle$ direction. The bright spot in the bottom of the images is a metal cluster that was deposited by controlled tip-sample contact for tip preparation and as a registry marker. Analysis of hundreds of such images led to the data set presented in this work.
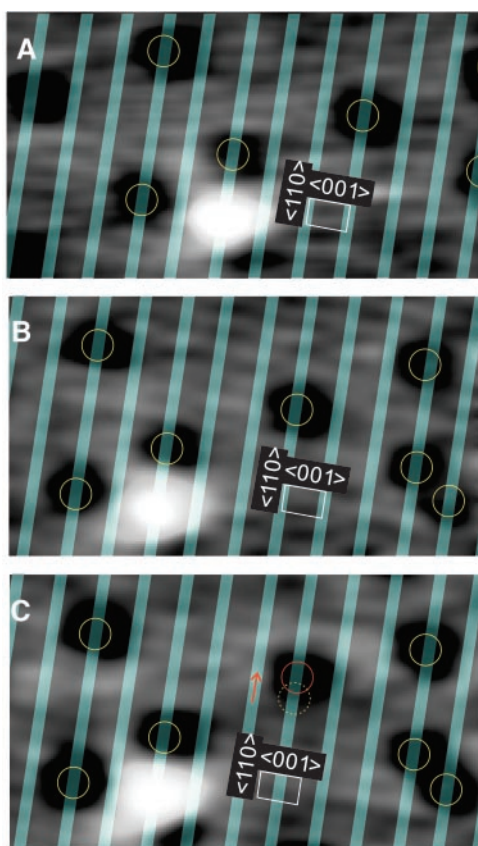
Fig. 3. (**A**) Schematic representation of Cu(110) indicating the crystallographic axis and interatomic distances. (**B**) Summary of experimental results of fs laser-induced diffusion and desorption for CO-Cu(110) measured with a total of ~3000 adsorbed molecules. The ratio of hopping across the close-packed rows ($\langle 001 \rangle$) to hopping along the rows ($\langle 1\bar{1}0 \rangle$) is 0.34.
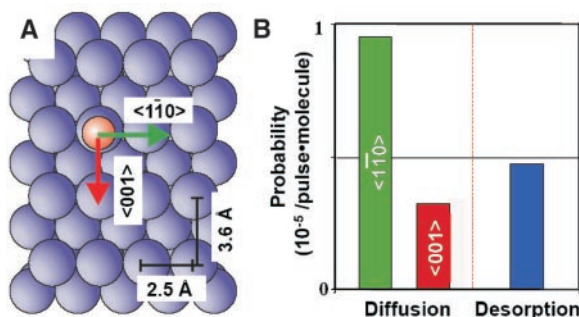
Fig. 4. (**A**) Calculated temperature profiles for the substrate electrons (T_{el}) and phonons (T_{ph}), and the adsorbate frustrated translational mode (T_{ad}) under laser irradiation of the Cu(110) sample. A peak electronic temperature $T_{el,max} = 3000$ K is reached, in contrast to a peak phonon temperature $T_{ph,max} = 140$ K. (**B**) The CO diffusion probability across the close-packed rows ($\langle 001 \rangle$) and along the rows ($\langle 1\bar{1}0 \rangle$) from the calculated increase in the adsorbate temperature.
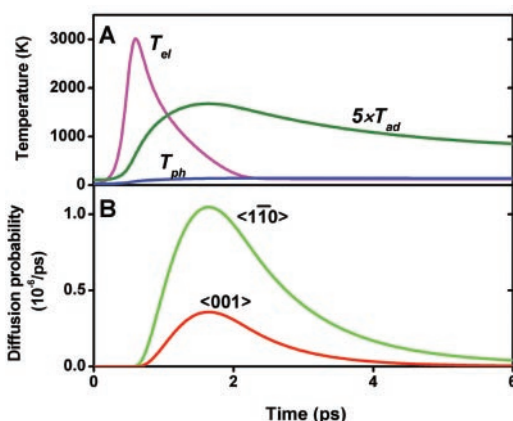
completeness both electronic and phononic couplings, the observed surface hopping is driven by the high electronic temperature.

The predicted temperature of the translational degree of freedom of the adsorbed CO molecules is shown in Fig. 4. We calculate a hopping probability per laser pulse of 2×10^{-6}. Given the severe approximations in this treatment, the agreement with the experimental result of 9×10^{-6} is satisfactory.

Can this analysis also explain the anomalous behavior with respect to adsorbate motion perpendicular to the atomic rows? Here, we cannot rely on an established Arrhenius rate law; all we know is that measurable hopping perpendicular to the atomic rows has not been observed under conventional thermal conditions. For purposes of a semiquantitative discussion, we assume that the lack of thermal diffusion in the perpendicular direction is attributable primarily to a larger energy barrier associated with the increased hopping distance to the next atom. Taking the lack of thermal diffusion across the rows in the measurements of Briner *et al.* (*18*) to mean that it is at least three orders of magnitude less probable than diffusion along the rows, we then deduced a minimum barrier height of 128 meV for diffusion across the rows, compared to a barrier of 97 meV for motion along the rows. With this parameter, our analysis of an electronically driven process yields a branching ratio for diffusion across the rows of 0.25. This value is comparable to our experimental finding of 0.34 but stands in sharp contrast to the equilibrium behavior. The observed branching ratio reflects the relative insensitivity to barrier heights of the electronically driven process.

These experiments show the possibility of combining direct imaging of molecules by STM with access to the ultrafast time scale by femtosecond laser excitation. The extension of these measurements to multiple-pulse laser excitation should retain the atomic-scale spatial resolution while providing further insight into the temporal evolution of ultrafast dynamical processes.

References and Notes
1. R. J. Hamers, D. G. Cahill, *Appl. Phys. Lett.* **57**, 2031 (1990).
2. S. Weiss, D. F. Ogletree, D. Botkin, M. Salmeron, D. S. Chemla, *Appl. Phys. Lett.* **63**, 2567 (1993).
3. V. Gerstner, A. Knoll, W. Pfeiffer, A. Thon, G. Gerber, *J. Appl. Phys.* **88**, 4851 (2000).
4. J. A. Prybyla, T. F. Heinz, J. A. Misewich, M. M. T. Loy, J. H. Glownia, *Phys. Rev. Lett.* **64**, 1537 (1990).
5. J. A. Prybyla, H. W. K. Tom, G. D. Aumiller, *Phys. Rev. Lett.* **68**, 503 (1992).
6. D. G. Busch, S. W. Gao, R. A. Pelak, M. F. Booth, W. Ho, *Phys. Rev. Lett.* **75**, 673 (1995).
7. L. M. Struck, L. J. Richter, S. A. Buntin, R. R. Cavanagh, J. C. Stephenson, *Phys. Rev. Lett.* **77**, 4576 (1996).
8. J. A. Misewich, S. Nakabayashi, P. Weigand, M. Wolf, T. F. Heinz, *Surf. Sci.* **363**, 204 (1996).
9. R. J. Finlay, T. H. Her, C. Wu, E. Mazur, *Chem. Phys. Lett.* **274**, 499 (1997).
10. G. Eichhorn, M. Richter, K. Al-Shamery, H. Zacharias, *Chem. Phys. Lett.* **289**, 367 (1998).
11. M. Bonn *et al.*, *Science* **285**, 1042 (1999).
12. D. N. Denzler, C. Frischkorn, C. Hess, M. Wolf, G. Ertl, *Phys. Rev. Lett.* **91**, 226102 (2003).
13. K. Watanabe, N. Takagi, Y. Matsumoto, *Phys. Rev. Lett.* **92**, 057401 (2004).
14. H. Petek, M. J. Weida, H. Nagano, S. Ogawa, *Science* **288**, 1402 (2000).
15. R. Gomer, *Rep. Prog. Phys.* **53**, 917 (1990).
16. T. T. Tsong, *Prog. Surf. Sci.* **67**, 235 (2001).
17. J. Ahner, J. T. Yates, *J. Chem. Phys.* **105**, 6553 (1996).
18. B. G. Briner, M. Doering, H.-P. Rust, A. M. Bradshaw, *Science* **278**, 257 (1997).
19. L. Bartels, M. Wolf, G. Meyer, K. H. Rieder, *Chem. Phys. Lett.* **291**, 573 (1998).
20. H. J. Lee, W. Ho, *Science* **286**, 1719 (1999).
21. A. J. Heinrich, C. P. Lutz, J. A. Gupta, D. M. Eigler, *Science* **298**, 1381 (2002); published online 24 October 2002; 10.1126/science.1076768.
22. T. Komeda, Y. Kim, M. Kawai, B. N. J. Persson, H. Ueba, *Science* **295**, 2055 (2002).
23. J. I. Pascual, N. Lorente, Z. Song, H. Conrad, H. P. Rust, *Nature* **423**, 525 (2003).
24. T. A. Germer, J. C. Stephenson, E. J. Heilweil, R. R. Cavanagh, *Phys. Rev. Lett.* **71**, 3327 (1993).
25. T. A. Germer, J. C. Stephenson, E. J. Heilweil, R. R. Cavanagh, *J. Chem. Phys.* **101**, 1704 (1994).
26. J. P. Culver, M. Li, Z. J. Sun, R. M. Hochstrasser, A. G. Yodh, *Chem. Phys.* **205**, 159 (1996).
27. J. P. Culver, M. Li, R. M. Hochstrasser, A. G. Yodh, *Surf. Sci.* **368**, 9 (1996).
28. M. Head-Gordon, J. C. Tully, *Phys. Rev. B* **46**, 1853 (1992).
29. C. Springer, M. Head-Gordon, *Chem. Phys.* **205**, 73 (1996).
30. J. T. Kindt, J. C. Tully, M. Head-Gordon, M. A. Gomez, *J. Chem. Phys.* **109**, 3629 (1998).
31. D. A. Micha, Z. G. Yi, *Chem. Phys. Lett.* **298**, 250 (1998).
32. Our experiments were performed with a variable-temperature STM, which was operated at a base pressure of 8×10^{-11} mbar. The Cu(110) sample was prepared by multiple cycles of sputtering (Ne^+ at 1 keV) and annealing (300 s at 900 K).
33. We gratefully acknowledge postdoctoral fellowships from the Humboldt Foundation, the Novartis Foundation, and the Academia Leopoldina. This work was supported by the U.S. Department of Energy, Office of Basic Energy Sciences, through Catalysis Science grant number DE-FG03-03ER15463/4; by the American Chemical Society Petroleum Research Fund; and by equipment grants F49620-96-1-0406 and F49620-01-1-0286 from the U.S. Air Force Office of Scientific Research.

Nanoparticles: Strained and Stiff

Benjamin Gilbert,[1] Feng Huang,[1] Hengzhong Zhang,[1] Glenn A. Waychunas,[2] Jillian F. Banfield[1,2]*

Nanoparticles may contain unusual forms of structural disorder that can substantially modify materials properties and thus cannot solely be considered as small pieces of bulk material. We have developed a method to quantify intermediate-range order in 3.4-nanometer-diameter zinc sulfide nanoparticles and show that structural coherence is lost over distances beyond 2 nanometers. The zinc-sulfur Einstein vibration frequency in the nanoparticles is substantially higher than that in the bulk zinc sulfide, implying structural stiffening. This cannot be explained by the observed 1% radial compression and must be primarily due to inhomogeneous internal strain caused by competing relaxations from an irregular surface. The methods developed here are generally applicable to the characterization of nanoscale solids, many of which may exhibit complex disorder and strain.

The electronic properties of nanoparticles can differ from those of their corresponding bulk form due to confinement effects caused only by their finite size, and because they are structurally distinct. Quantum confinement is the dominant size effect, and has been well studied (*1–6*). Structural deviations in nanoparticles relative to bulk material are not well understood because they are hard to resolve experimentally (*7*). Consequently, theoretical models of nanoparticles generally assume they have bulklike interior structure (*8*). Tight-binding calculations that optimized nanoparticle structure, and assumed full theoretical passivation of surface anions, suggested that surfaces of nanoparticles relax in a manner comparable to that of bulk surfaces (*9*). However, classical and quantum molecular dynamics simulations have indicated that disorder may pervade throughout nanoparticles (*10*, *11*). We previously showed that nanoparticles can undergo substantial transformation in structure at low temperature, driven by surface interactions, indicating that internal strain depends upon the nature of the surroundings as well as size (*12*). However, a detailed description of the strain within nanoparticles has not been experimentally obtained. We combine pair distribution function and extended x-ray absorption fine structure (EXAFS) analyses to quantify the structural distortion within mercaptoethanol-coated zinc sulfide (ZnS) nanoparticles and the consequent changes in lattice dynamics.

[1]Department of Earth and Planetary Sciences, University of California at Berkeley, Berkeley, CA 94720, USA. [2]Earth Sciences Division, Lawrence Berkeley National Lab, One Cyclotron Road, Berkeley, CA 94720, USA.

*To whom correspondence should be addressed. E-mail: jill@eps.berkeley.edu

Mercaptoethanol-coated ZnS nanoparticles were synthesized using the method of Vogel *et al.* (*13*). Fits to small-angle x-ray scattering (SAXS) data indicate that the average diameter is 3.4 nm with a full width at half maximum (FWHM) of 0.6 nm for the size distribution (Fig. 1), in close agreement with the size estimated from ultraviolet-visible absorption spectroscopy (fig. S1) (*14*). High-resolution transmission electron microscope (HRTEM) data confirm the size determined by SAXS and show that the nanoparticles are approximately spherical (fig. S2).

We acquired wide-angle x-ray scattering (WAXS) patterns from powders of bulk ZnS (sphalerite structure) and mercaptoethanol-coated ZnS nanoparticles (fig. S3). Although the WAXS patterns from the nanoparticles and from the bulk material have peaks in essentially the same positions (indicating that they share the same basic structure), the peaks from the nanoparticles are wider and less intense. We obtained the real-space interatomic distance correlation functions, commonly known as pair distribution functions (PDFs), from the WAXS data using standard methods (*14–17*). The PDFs of bulk ZnS and ZnS nanoparticles are given in Fig. 2A. The PDF reveals interatomic correlations at much greater distances than can be achieved by EXAFS (see below) (*18–20*). In contrast to Bragg peak analysis of diffraction data (*21*), the PDF additionally uses information from total diffuse x-ray scattering.

The nanoparticle PDF approaches zero at 2.0 nm rather than at 3.4 nm, indicating the presence of structural disorder. A detailed analysis of the disorder in the nanoparticles can be made by comparing the observed PDF to one obtained by truncating the PDF from bulk ZnS with a real-space curve associated with a 3.4-nm-diameter sphere (*14, 22*). This operation creates the PDF of "ideal" ZnS sphalerite nanoparticles, also shown in Fig. 2A, allowing finite particle-size effects on the diffraction data to be separated from structural effects.
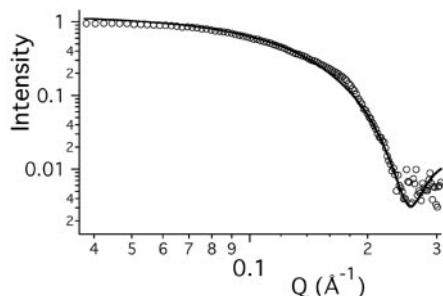
The PDF for real ZnS nanoparticles is distinct from that of ideal nanoparticles in the following respects: (i) The first-shell PDF peak intensity, representing Zn-S bonds throughout the nanoparticles, is lower for real than for ideal nanoparticles. (ii) PDF peak intensities at higher correlation distances diminish more rapidly in the real nanoparticles. (iii) PDF peak widths are broader in the real nanoparticles. (iv) PDF peak positions are shifted. Observations (i) to (iii) indicate the presence of two distinct forms of excess static disorder within the real nanoparticles relative to bulk sphalerite (Fig. 3, A and B). The low dependence on temperature of WAXS measurements (fig. S7) indicates that this excess disorder is structural (i.e., static disorder) rather than vibrational (i.e., thermal disorder). The reduction and broadening of all peaks in the PDF indicate that the random mean squared displacements about equilibrium positions (uncorrelated with the positions of neighboring atoms) are increased (Fig. 3A). However, this type of disorder broadens all peaks in the PDF equivalently and cannot

explain the loss of intensity with increasing interatomic distance. Thus, a second effect is that strain within the nanoparticle causes correlated shifts in the equilibrium positions themselves, relative to the perfect sphalerite lattice (Fig. 3B). This effect is cumulative with increasing interatomic distance and causes peak intensities in the real nanoparticle PDF to vanish at a correlation length of about 2.0 nm.

In addition, observation (iv) indicates a contraction of nanoparticle bond lengths (Fig. 3C). We quantified the lattice contraction and the two types of disorder by fitting real-space interatomic correlations from the sphalerite structure to the experimental PDFs (*14*). From a fit to the bulk sphalerite data (fig. S5), we obtained the thermal contribution to PDF peak broadening by fitting the mean squared relative displacement (MSRD) while accounting empirically for correlated atomic motion (*23*). From a fit to the nanoparticle data that accounted for the nanoparticle size and incorporated the thermal MSRD value obtained from the bulk, we obtained two

Fig. 1. Small-angle x-ray scattering (SAXS) data (circles) from mercaptoethanol-coated ZnS nanoparticles in aqueous solution. The fit (solid line) is for polydisperse dilute spheres of average diameter 3.4 nm and Schultz size distribution FWHM of 0.6 nm.
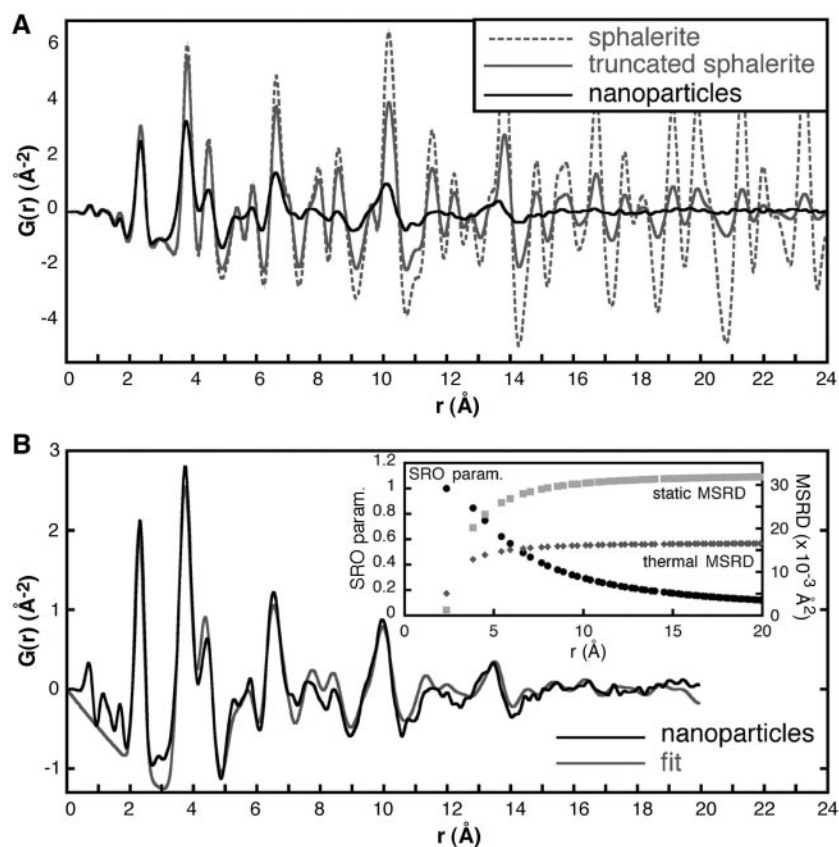
Fig. 2. (**A**) PDFs of bulk ZnS (sphalerite, dashed gray curve), bulk ZnS following truncation by the shape factor for a 3.4-nm-diameter sphere (gray curve), and mercaptoethanol-coated ZnS nanoparticles of average diameter 3.4 nm (black curve). (**B**) Theoretical fit (gray curve) to the PDF of ZnS nanoparticles (black curve), accounting for finite particle size and including the thermal mean square relative displacement (MSRD) obtained from a fit to the bulk sphalerite data (inset, diamonds). In addition, the fit includes a static harmonic disorder contribution to the MSRD (inset, squares) and a short-range order (SRO) parameter (inset, circles). The MSRD contributions broaden the PDF peaks; the SRO parameter is an intensity scaling factor.

additional static disorder parameters that have a dependence on interatomic distance, r. The first parameter is the static MSRD contribution to PDF peak broadening due to random displacements about equilibrium positions. The second is a short-range order parameter, $P^{SRO}(r)$, where $0 < P^{SRO} \leq 1$, that causes the loss of PDF peak intensity due to the correlated shifts in equilibrium positions. The best-fit results are given in Fig. 2B.

The static MSRD contribution is shown in the inset to Fig. 2B. Its form resembles that of the thermal MSRD (obtained from bulk ZnS) in that there is less disorder at sh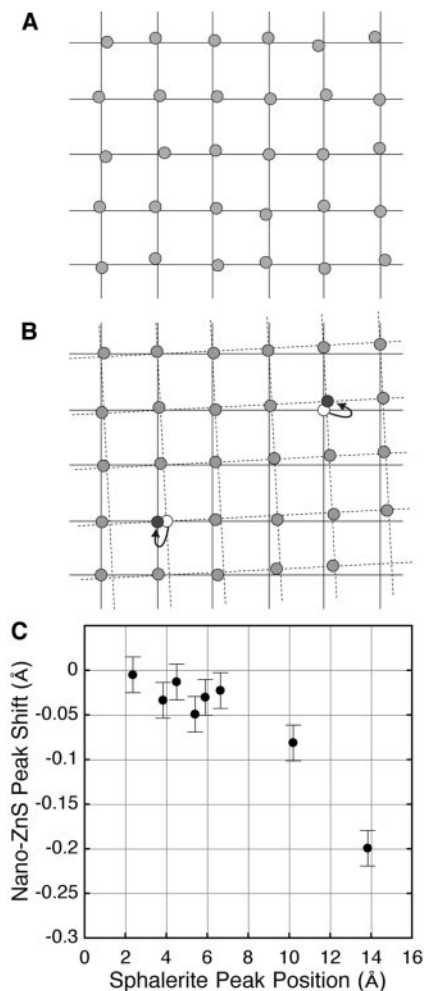ort correlation lengths. For thermal disorder, this line-shape is due to correlated atomic motion (23). By analogy, the static disorder is also correlated over short distances.

The best-fit curve $P^{SRO}(r)$ approaches zero with increasing correlation distance. This shows that, at larger interatomic distances, fewer atoms maintain definite structural relationships relative to other atoms in the nanoparticle. This finding is consistent with preservation of short-range order, strong reduction in intermediate-range order, and complete loss of structural coherence at distances greater than 2 nm (16).

From the PDF fit, we measured a mean 1% radial contraction of interatomic correlations in the real nanoparticles. However, the peak shifts are more complex, as indicated by a plot of nanoparticle peak shift versus sphalerite peak position (Fig. 3C). From studies on a model nanoparticle subjected to structural modifications, we can show that simple models for internal strain can be distinguished by the shifts in the associated PDF peak positions (fig. S8). However, we conclude that the strain within the real nanoparticles is more complex than that described by any of the simple models that we considered, including uniform, or surface-weighted, radial strain; linear strain; or the presence of stacking faults.

The loss of structural coherence and the inability of simple strain models to describe peak shifts in PDF data point to the existence of complex strain fields within the nanoparticles. We infer that the pervasive internal distortion is due to strain that arises from the competing attempts of the structurally diverse terminating surfaces that encompass the nanoparticles to adopt lower energy configurations. Even with strong chemical passivation, nanoparticle surfaces are generally nonstoichiometric [because capping ligands interact with surface cations or anions only, e.g. (20)] and incompletely capped [due to steric limits, e.g. (24)]. The inherent limit of surface passivation creates structural effects throughout the nanoparticle interior, despite the highly crystalline appearance within the transmission electron micrographs (fig. S2) or from x-ray diffraction (fig. S3). We anticipate that the observed forms of disorder are general features of nanoscale solids and will be especially pronounced where there is minimal passivation.

Because internal strain and disorder can substantially affect materials properties, we investigated the lattice dynamics of the ZnS samples. We obtained Zn K-edge EXAFS spectra between 3 and 500 K from the same bulk ZnS and ZnS nanoparticles used in the PDF analysis (Fig. 4A and fig. S9). The EXAFS spectra from bulk ZnS contain peaks from long-range structure that are weak or absent in the spectra from the nanoparticles, even at low temperature.

We obtained the first-shell (Zn-S) MSRD as a function of temperature by fitting the data in Fig. 4A to ZnS scattering paths obtained through simulation using the FEFF code (14, 25). In Fig. 4B, the nanoparticle MSRD curve is offset from that of the bulk, indicating the presence of excess static disorder in the nanoparticles. The room-temperature EXAFS and PDF analyses yield consistent measurements of this excess structural disorder. First-shell combined thermal and static MSRDs in bulk and nanoparticles are 5.1×10^{-3} $Å^2$ and 6.4×10^{-3} $Å^2$, respectively, from fits to the EXAFS data, and 5.0×10^{-3} $Å^2$ and 6.6×10^{-3} $Å^2$ from fits to the PDF data.

From Fig. 4B, the amplitudes of vibrations in bulk ZnS begin to increase rapidly with temperature at $\sim$70 K, but this increase is not seen in the nanoparticles until $\sim$130 K. This indicates structural stiffening in the nanoparticles.

We fitted the temperature response of the MSRD to a theoretical expression for the anharmonic Einstein oscillator (14, 26). In the Einstein model, the Zn-S bonding pair is considered to vibrate in the center of mass

Fig. 3. Structural modifications in ZnS nanoparticles can be decomposed into two distinct disorder types, plus lattice contraction. (**A**) Random displacement disorder. Individual atoms are randomly displaced from sites on a single lattice. (**B**) Strain-driven distortion. Local structure is maintained, but at larger interatomic distances, atoms lie farther from the positions expected for an undistorted lattice. (**C**) Lattice contraction. The shift in PDF peak positions observed in ZnS nanoparticles is plotted against the positions of the equivalent peaks in the bulk sphalerite PDF.

Fig. 4. (**A**) Fourier transform magnitude of k^3-weighted Zn K-edge EXAFS spectra of bulk ZnS (sphalerite) and ZnS nanoparticles as a function of temperature. (**B**) Temperature dependence of the first-shell mean square relative displacement (MSRD) obtained from fits to the EXAFS data. The lines are fits to the first-shell MSRD data using an anharmonic Einstein oscillator model to extract characteristic vibration frequencies for the two samples. A discussion of the errors in the MSRD determination is given in the Supporting Online Material (14).

frame under the influence of the interatomic potential of the remaining stationary lattice. The characteristic vibration frequency obtained for bulk ZnS is 7.12 ± 1.2 THz. This value is consistent with that reported previously for bulk CdSe (27, 28). In comparison, the characteristic vibration frequency for the ZnS nanoparticles is 11.6 ± 0.4 THz. Although these values cannot be simply related to the full phonon dispersion relations for ZnS, the trend clearly indicates lattice stiffening. This implies that the heat capacity of the ZnS lattice, excluding surface-bound species, is lower in nanoparticles than in equivalent bulk material.

There are several features of the nanoparticles that may be responsible for the lattice stiffening. First, the PDF analysis indicates ~1% bond length compression. Because crystalline materials show an increase in vibration frequencies with pressure (away from phase transitions), the compression will cause lattice stiffening. However, compression of this magnitude alone cannot explain the large increase in vibration frequency. The average change in bond length provides an incomplete picture because disorder within the nanoparticles leads to a large distribution of bond lengths, and all deviations from the equilibrium bond lengths may increase the vibration frequency. We thus find that structural disorder is principally responsible for the lattice stiffening.

In summary, we quantitatively determined nanoparticle crystallinity and disorder, using a PDF-based method. We found that the best description of nanoparticle structure includes bond length contraction, random disorder, and a type of disorder characterized by correlated atomic displacements. The internal disorder is observed despite the presence of strongly bound surface ligands. We conclude that even with strong chemical passivation, surface atoms exist in diverse unsatisfied bonding environments at the surface that drive inhomogeneous internal strain. A striking consequence of the internal disorder is that the nanoparticles are much stiffer than expected from the measured bond length contraction. Our approach is an important step toward a realistic description of nanoparticle structure that includes internal strain, which is likely to be a general feature of nanoscale solids. Because lattice contraction and disorder will separately modify electronic properties, inclusion of these effects is essential for accurate nanoparticle calculations.

References and Notes

1. R. Rossetti, R. Hull, J. M. Gibson, L. E. Brus, *J. Chem. Phys.* **82**, 552 (1985).
2. C. B. Murray, D. J. Norris, M. G. Bawendi, *J. Am. Chem Soc.* **115**, 8706 (1993).
3. A. P. Alivisatos, *Science* **271**, 933 (1996).
4. L. E. Brus, *J. Chem. Phys.* **80**, 4403 (1984).
5. P. E. Lippens, M. Lannoo. *Phys. Rev. B* **39**, 10935 (1989).
6. A. Franceschetti, A. Zunger, *Phys. Rev. Lett.* **78**, 915 (1997).
7. B. Palosz et al., *Z. Kristallogr.* **217**, 497 (2002).
8. L.-W. Wang, A. Zunger, *Phys. Rev. B* **53**, 9579 (1996).
9. K. Leung, K. B. Whaley, *J. Chem. Phys.* **110**, 11012 (1999).
10. E. Rabani, *J. Chem. Phys.* **115**, 1493 (2001).
11. A. Puzder, A. J. Williamson, F. A. Reboredo, G. Galli, *Phys. Rev. Lett.* **91**, 157405 (2003).
12. H. Zhang, B. Gilbert, F. Huang, J. F. Banfield. *Nature* **424**, 1025 (2003).
13. W. Vogel, P. H. Borse, N. Deshmukh, S. K. Kulkarni, *Langmuir* **16**, 2032 (2000).
14. See supporting online material.
15. B. J. Thijsse, *J. Appl. Crystallogr.* **17**, 61 (1984).
16. A. C. Wright et al., *J. Non-Cryst. Solids* **129**, 213 (1991).
17. S. J. L Billinge, M. F. Thorpe, Eds., *Local Structure from Diffraction* (Plenum, New York, 1998).
18. J. Rockenburger, et al. *J. Chem. Phys.* **108**, 7807 (1998).
19. M. A. Marcus, W. Flood, M. Steigerwald, L. Brus, M. Bawendi. *J. Phys. Chem.* **95**, 1572 (1991).
20. A. C. Carter, et al., *Phys. Rev. B* **55**, 13822 (1997).
21. M. G. Bawendi, A. R. Kortan, M. L. Seigerwald, L. E. Brus, *J. Chem. Phys.* **91**, 7282 (1989).
22. G. Mason, *Nature* **217**, 733 (1968).
23. I.-K. Jeong, Th. Proffen, F. Mohiuddin-Jacobs, S. J. L. Billinge, *J. Phys. Chem. A* **103**, 921 (1999).
24. J.-J. Shiang, A. V. Kadavanich, R. K. Grubbs, A. P. Alivisatos, A. P. 1995. *J. Phys. Chem.* **99**, 17417 (1995).
25. A. L. Ankudinov, B. Ravel, J. J. Rehr, S. D. Conradson, *Phys. Rev. B* **58**, 7565 (1998).
26. A. I. Frenkel, J. J. Rehr, *Phys. Rev B* **48**, 585 (1993).
27. G. Dalba, et al., *Phys. Rev. B* **58**, 4793 (1998).
28. For two atoms vibrating in their center of mass frame in an effective pair potential characterized by spring constant k, the bond frequency, $\omega = \sqrt{k/\mu}$, where μ is the reduced mass. For CdSe, $1/\mu = 1/M_{Cd} + 1/M_{Se}$ (M_{Cd} is the atomic mass of Cd). For CdSe, $\nu_{CdSe} = 4.97$ THz (27). For ZnS, we obtained $\nu_{ZnS} = 7.12$ THz. The interatomic force constants for CdSe and ZnS are expected to be similar. If we assume that $k_{CdSe} = k_{ZnS}$, then $\left.\frac{\omega_{CdSe}}{\omega_{ZnS}}\right|_{predicted} = \sqrt{\frac{\mu_{ZnS}}{\mu_{CdSe}}} = 0.68$. In fact, $\left.\frac{\omega_{CdSe}}{\omega_{ZnS}}\right|_{observed} = \sqrt{\frac{\mu_{ZnS}}{\mu_{CdSe}}} = 0.676$. Hence, the results presented here for bulk ZnS and in (27) for bulk CdSe are in good agreement.
29. We thank M. Toney for discussion on PDF analysis, C. Kim for practical advice on EXAFS acquisition, and M. Zach for assistance with sample holder fabrication. SAXS data were acquired on beamline 1.4 at the Stanford Synchrotron Radiation Lab (SSRL), and we thank J. Pople. Zn K-edge EXAFS were acquired on beamline 4-3 at SSRL, and we thank M. Lattimer. WAXS data were acquired on beamline 11-ID-C at the Advanced Photon Source (APS), Argonne National Laboratory, and we thank Y. Ren and M. Beno. The U.S. Department of Energy, Office of Basic Energy Sciences, supports use of the APS (Contract No.W-31-109-Eng-38) and the SSRL (Contract No. DE-AC03-76SF00515). HRTEM was performed at the National Center for Electron Microscopy, Berkeley, CA, and we thank C. Song. Financial support for this work came from the U.S. Department of Energy, the National Science Foundation, and Lawrence Berkeley National Laboratory.

Grain Boundary–Mediated Plasticity in Nanocrystalline Nickel

Zhiwei Shan,[1] E. A. Stach,[2] J. M. K. Wiezorek,[3] J. A. Knapp,[4] D. M. Follstaedt,[4] S. X. Mao[1]*

The plastic behavior of crystalline materials is mainly controlled by the nucleation and motion of lattice dislocations. We report in situ dynamic transmission electron microscope observations of nanocrystalline nickel films with an average grain size of about 10 nanometers, which show that grain boundary–mediated processes have become a prominent deformation mode. Additionally, trapped lattice dislocations are observed in individual grains following deformation. This change in the deformation mode arises from the grain size–dependent competition between the deformation controlled by nucleation and motion of dislocations and the deformation controlled by diffusion-assisted grain boundary processes.

The plastic deformation of coarse-grained metals at relatively low temperature is usually mediated by the nucleation and motion of

[1]Department of Mechanical Engineering, University of Pittsburgh, 648 Benedum Hall, Pittsburgh, PA 15261, USA. [2]National Center for Electron Microscopy (NCEM), Lawrence Berkeley National Laboratory (LBNL), Berkeley, CA 94720, USA. [3]Department of Materials Science and Engineering, University of Pittsburgh, Pittsburgh, PA 15261, USA. [4]Physical and Chemical Science Center, Sandia National Laboratories (SNL), Albuquerque, NM 87185, USA.

*To whom correspondence should be addressed. E-mail: smao@engr.pitt.edu

dislocations. However, for nanocrystalline metals, dislocation sources and pile-ups are not expected to exist within the individual grains because of the limited grain size and the much higher fraction of grain boundary (GB) atoms (1–8). It has been proposed (3–6) that GB-mediated plasticity (e.g., GB sliding and/or grain rotation) substitutes for conventional dislocation nucleation and motion as the dominant deformation mechanism when grain sizes are reduced below a certain value. This conjecture has been supported by molecular dynamics simulations (5, 7, 8) and indirect experimental evidence (2, 9). Dy-

namic in situ deformation studies (*10–13*) have been performed in the transmission electron microscope (TEM) to observe the deformation mechanisms that are active in nanocrystalline metals with grain sizes in the regime where a transition in mechanism is predicted. However, direct experimental evidence of a transition in deformation mechanisms, i.e., from a dislocation-mediated deformation to a GB-mediated deformation, has not been reported (*10–14*).

The Ni samples used in this study were synthesized by directing a high-energy pulsed KrF excimer laser onto a high-purity Ni target under vacuum with a base pressure of about 2×10^{-7} torr. The resulting Ni plasma was deposited onto a [001] NaCl substrate with a nominal thickness of 150 nm. TEM observations (Fig. 1A) indicate that the as-deposited Ni consists of roughly equiaxed grains with random orientations (inset in Fig. 1A). Statistical measurements of dark-field TEM (DFTEM) images reveal a narrow, log-normal grain size distribution, ranging from several nanometers to 23 nm with an average value (±SD) of 9.7 ± 3.9 nm (Fig. 1B). High-resolution TEM (HRTEM) shows that most grains are separated by large-angle GBs. Additional GB phases, porosity, or intergranular microcracks were not detected in these samples after deposition.

The nanocrystalline nickel films were deformed by in situ TEM tensile straining (*15*). Figure 2A is a typical DFTEM image of a Ni sample in the undeformed state. Figure 2B is the corresponding selected area diffraction pattern (SADP) before deformation. Upon straining, bright-field TEM observations showed rapid changes in contrast that occur continuously in many different grains. This has previously been identified as evidence of dislocation activity (*11, 13*). However, our in situ DFTEM observations revealed that the deformation behavior is substantially more complex. We found many strongly diffracting features that are much larger than the initial average grain size and are formed immediately in the deformed zone upon straining (e.g., the feature marked by a white arrow is about 60 nm in diameter in Fig. 2C). The SADP from the deformed area (Fig. 2D, taken from the same region and with the same selected area aperture size as the area in Fig. 2B) exhibited fewer, asymmetrically distributed diffraction spots in each of the diffraction rings in comparison with the SADP obtained before deformation (Fig. 2B). The reduced background intensity and improved contrast in the SADP of the deformed area (Fig. 2D) indicate that the sample thinned locally during deformation. Higher magnification DFTEM micrographs revealed that the large bright features observed after deformation consist of a num-

ber of smaller grains (e.g., Fig. 2E) rather than a single large grain resulting from stress-assisted grain growth. A comparison of the undeformed state (Fig. 2, A and B) and the postdeformation state (Fig. 2, C, D, and E) indicates that groups of neighboring grains underwent an orientation change during straining and formed numerous grain agglomerates. Because of the nature of the image formation mechanism of DFTEM images (*15*), we can conclude that the smaller grains in these agglomerates exhibit essentially edge-on orientations of their {111} and/or {200} lattice planes. The schematic in Fig. 2F depicts a possible crystallographic substructure associated with an agglomerated group of grains that would be consistent with the type of contrast observed in the DFTEM micrographs after straining (Fig. 2E).

To elucidate further the mechanism responsible for the formation of the agglomerated grain regions, real-time observations were performed. Contrast changes in areas subject to high strain during in situ TEM straining experiments were recorded in the dark-field (DF) mode. The DF micrographs
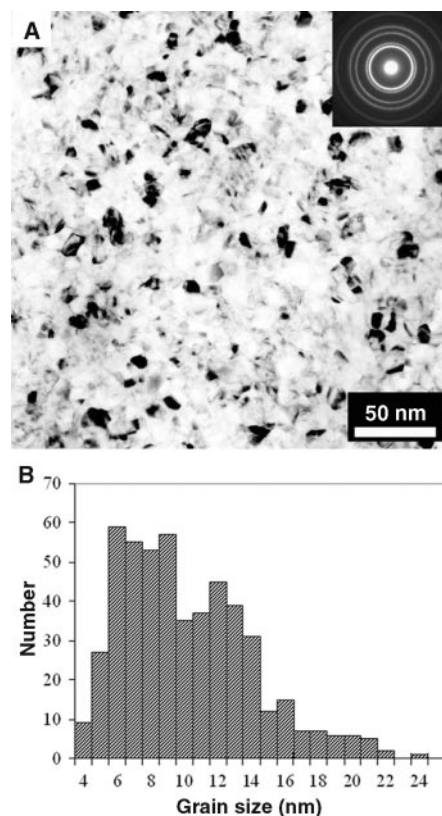

Fig. 1. TEM observations of the typical microstructure in the as-deposited nanocrystalline nickel films. The bright-field TEM micrograph (**A**) and the SADP [inset in (A)] show roughly equiaxed grains with random orientations. The statistical distributions for grain size (**B**) were obtained from multiple TEM images of the same sample.

shown in Fig. 3 are still frames extracted from a typical dynamic sequence of images taken during the application of a single displacement pulse. The times listed on each still frame are based on the video-acquisition rate of 30 frames per second. At the beginning of this sequence, there were no grains in a strongly diffracting condition in the area indicated by the white arrow (Fig. 3A). After 1/30 of a second, a bright spot emerged from a grain about 6 nm in diameter and remained well defined in size as a single, approximately equiaxed grain until $t = 0.1$ s (Fig. 3B). Over the next couple of frames, a number of additional neighboring grains rotated into strongly diffracting conditions for either the {200} or {111} planes. Other nearby grains, which were in a strong diffraction condition, did not rotate out of contrast, confirming that there had been no global rotation of the specimen area and that the rotations observed at
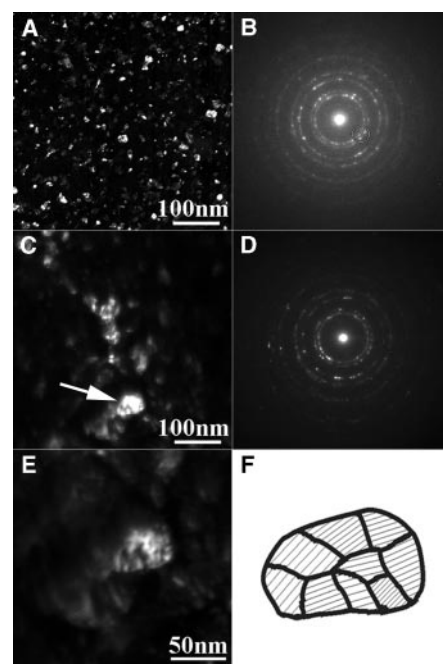
Fig. 2. TEM micrographs showing the evolution of the Ni microstructure during in situ straining. (**A**) DFTEM image of as-deposited nanocrystalline Ni film before deformation. Part of the {111} and {200} diffraction rings were selected as the image-forming diffraction vectors, as indicated schematically by the white circle in (B) (all of the DF images here are taken with this condition). (**B**) Corresponding SADP of undeformed Ni film. (**C**) DFTEM micrograph of as-deposited Ni after deformation. Features in strong diffraction contrast much larger than the average grain size are discernible (e.g., white arrow). (**D**) The corresponding SADP of the deformed Ni film. (**E**) Higher magnification DFTEM image of the feature marked by the white arrow in (C). (**F**) Schematic sketch of a possible multigrain substructure of the large features in bright contrast observed after deformation [e.g., in (C)]. The lines represent the trace of those edge-on planes.

the arrowed location were internal changes of the sample structure. Additionally, the small "notch" discernible at the lower left corner of the agglomerate (Fig. 3, D and E) indicates that the growth in size of this agglomerated group of grains was not isotropic and involved sudden rotation of individual grains. At $t = 0.5$ s, the group of grains had grown into an elongated equiaxed shape with an approximate size of 60 nm along the short axis and 80 nm along the long axis. After this very rapid morphological change, the rate of growth of the grain agglomerate decreased significantly.

These TEM observations confirm the prediction that GB-mediated deformation can become prominent when the average grain size of a material decreases below some critical value. Several theoretical studies have considered the operation of grain rotation and GB sliding as possible deformation modes (16–18). These studies assume that the driving force for grain rotation is the net torque on a grain, which results from the misorientation dependence of the energy of the GBs that delineate a given grain from its neighbors. Grain rotation is viewed as a sliding problem on the periphery of the grain; changes in the grain shape during rotation are assumed to be accommodated by diffusion, through either the GBs or the grain interiors. If GB diffusion is dominant (a reasonable assumption in nanocrystalline metals with grain size below a certain value at room temperature), it would yield a d^{-4} dependence on the rotation rate, where d is the grain size. This is a marked dependence on grain size. For example, when d is 60 and 6 nm, respectively, and all other factors are identical, the grain rotation rate for the latter will be 10^4 higher than for the former. This qualitatively explains our observation of extremely rapid formation of the grain agglomerates by rotation during straining. The rate of the grain rotation processes will generally decrease with time, as a state of new equilibrium is approached locally.

Previous TEM observations have indicated that dislocations may still play a role in deformation even if the average grain size falls below the grain size at which GB processes may become active (13). Hence, it may be expected that evidence of lattice dislocation glide activity during straining would exist in the microstructure of deformed nanostructured metals. However, detection of such evidence by postmortem TEM investigations, for instance in the form of dislocation debris or trapped lattice dislocations, has proven unsuccessful (19). Yamakov et al. (20) suggested that trapped dislocations may only exist inside a nanograin under very high external or internal stress and that the removal of the stress can lead to reversible reabsorption of the partial dislocations at the presumed GB source. However, the traditional sample preparation methods, such as mechanical thinning followed by ion thinning or twin jet polishing will inevitably result in some relaxation of a previously deformed microstructure. This makes it difficult to verify the assumption, and therefore limits further understanding of this important fundamental process. During this study, HRTEM images of suitably oriented grains that were still under stress were obtained in the thin area that was produced by the deformation.
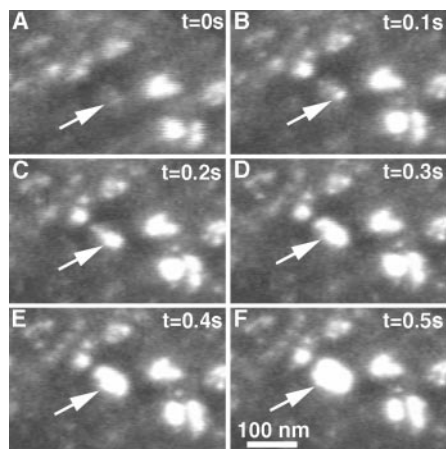
Trapped lattice dislocations were detected in some of the grains. Figure 4 shows an example of an HRTEM micrograph of one of these trapped dislocations (white T) trapped in the vicinity of a GB (delineated by the black dashed line). The inset at the upper right corner in Fig. 4 is an inverse Fourier-filtered image of the region framed by the white box, which displays the dislocation more clearly. Local Burgers circuits were employed to determine the Burgers vectors of the dislocations in HRTEM. As for Fig. 4, the Burgers vector of the trapped dislocation was determined to be consistent with $b = \frac{a}{2} <110>$,

where a is the lattice constant of Ni—i.e., a unit dislocation of the face-centered cubic lattice on a $(\bar{1}11)$ plane (15). The frequent observation of trapped lattice dislocations in the postdeformation state indicates that dislocation-mediated deformation is still active even when the average grain size is about 10 nm. Two other recent experiments have also detected the presence of dislocations in nanocrystalline Ni while under stress (21, 22), although for larger grains (20 and 26 nm, respectively).

The physics governing the observed deformation crossover can be understood by considering the effect of grain size on the different operative processes. A number of computational simulations have predicted that GBs can act as dislocation sources in nanocrystalline materials (5, 23, 24). Based on these simulations, Chen et al. (25) proposed a dislocation-based model suggesting that the nucleation stress for both perfect and partial dislocations is inversely proportional to the grain size. Furthermore, this model predicts the existence of a critical grain size d_c, below which the deformation mechanism will change from one controlled by normal unit dislocation motion to one controlled by partial dislocation activity. If we take the dislocation core parameter $\alpha = 1$, the shear modulus as $\mu_{Ni} = 95$ GPa (26), and the stacking fault energy of nickel as ~0.128 to 0.24 J/m^2 (27, 28), the critical size (d_c) for Ni then ranges from ~11 to 22 nm. However, according to this theory, the lower bound of the nucleation stress for partial dislocations (for $d = 23$ nm) in our sample is as high as ~2.1 to 2.8 GPa. This indicates that very high local stress is necessary for the nucleation of partial dislocations in nanocrystalline Ni with an average grain size of 10 nm. In contrast, the rate of GB-mediated deformation, as mentioned above, increases rapidly with a scaling of d^{-4}. Thus, the deformation mechanism crossover is an inevitable result of the competition between the deformation controlled by nucleation and motion of dislocations (unit and partial) and the deformation controlled by GB-related deformation accommodated mainly by GB diffusion with decreasing grain size.

In a critical part of our investigation, we

Fig. 3. DFTEM observation of the rapid genesis of an agglomerate (e.g., white arrow) depicted by individual still frames extracted from a dynamic video sequence. (**A**) At $t = 0$ s, no grains in the strong diffraction condition are near the white arrow. (**B**) At $t = 0.1$ s, a grain in the strong diffraction condition with a size of about 6 nm is visible. (**C**) At $t = 0.2$ s, a group of grains in bright contrast with a size of about 28 nm is visible. (**D**) At $t = 0.3$ s, the group of grains has a nearly elliptical shape, with dimensions of 60 by 35 nm. (**E**) At $t = 0.4$ s and (**F**) $t = 0.5$ s, the size of the group of grains increases to maximum dimensions of about 80 by 60 nm.
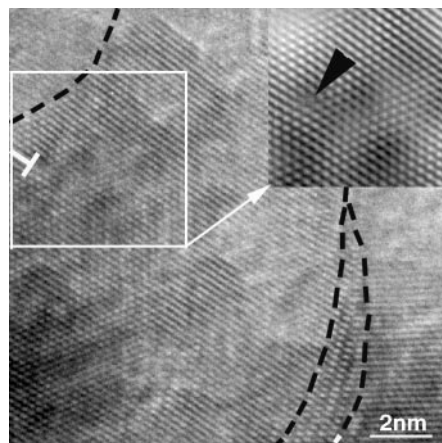

Fig. 4. A typical HRTEM image of a thin area formed by deformation. A dislocation (white T) is trapped inside a grain close to the GB (delineated by black dashed line). The inverse Fourier-filtered image (**inset**) from inside the white box shows the dislocation (black arrowhead) with more clarity.

observed the rapid sequence of initial grain realignments by closely examining successive video frames in DFTEM mode. The frequent observation of the GB-mediated deformation reported here (Figs. 2 and 3) would not have been possible in bright-field TEM conditions because of the inherent difficulty with differentiating the contrast changes caused by GB-related deformations from those caused by the motion of lattice dislocations in small grains (e.g., less than 20 nm).

References and Notes

1. J. R. Weertman et al., Mater. Res. Soc. Bull. 24, 44 (1999).
2. C. C. Koch, D. G. Morris, K. Lu, A. Inoue, Mater. Res. Soc. Bull. 24, 54 (1999).
3. H. Van Swygenhoven, Science 296, 66 (2002).
4. J. R. Weertman, in Nanostructured Materials: Processing, Properties and Applications, C. C. Koch, Ed. (William Andrews Publishing, Norwich, New York, 2002), pp. 397–421.
5. J. Schiotz, F. D. DiTolla, K. W. Jacobsen, Nature 391, 561 (1998).
6. S. Yip, Nature 391, 532 (1998).
7. J. Schiotz, K. W. Jacobsen, Science 301, 1357 (2003).
8. V. Yamakov, D. Wolf, S. R. Phillpot, A. K. Mukherjee, H. Gleiter, Nature Mater. 3, 43 (2004).
9. C. A. Schuh, T. G. Nieh, T. Yamasaki, Scripta Mater. 46, 735 (2002).
10. M. Ke, S. A. Hackney, W. W. Milligan, E. C. Aifantis, Nanostructured Mater. 5, 689 (1995).
11. C. J. Youngdahl, J. R. Weertman, R. C. Hugo, H. H. Kung, Scripta Mater. 44, 1475 (2001).
12. K. S. Kumar, S. Suresh, M. F. Chisholm, J. A. Horton, P. Wang, Acta Mater. 51, 387 (2003).
13. R. C. Hugo et al., Acta Mater. 51, 1937 (2003).
14. H. V. Swygenhoven, J. R. Weertman, Scripta Mater. 49, 625 (2003).
15. Materials and methods are available as supporting material on Science Online.
16. R. Raj, M. F. Ashby, Metal. Trans. 2, 1113 (1971).
17. K. E. Harris, V. V. Singh, A. H. King, Acta Mater. 46, 2623 (1998).
18. D. Moldovan, D. Wolf, S. R. Phillpot, Acta Mater. 49, 3521 (2001).
19. M. Legros, B. R. Elliott, M. N. Rittner, J. R. Weertman, K. J. Hemker, Philos. Mag. A 80, 1017 (2000).
20. V. Yamakov, D. Wolf, M. Salazar, S. R. Phillpot, H. Gleiter, Acta Mater. 49, 2713 (2001).
21. R. Mitra, W.-A. Chiou, J. R. Weertman, J. Mater. Res. 19, 1029 (2004).
22. Z. Budrovic, H. Van Swygenhoven, P. M. Derlet, S. V. Petegem, B. Schmitt, Science 304, 273 (2004).
23. V. Yamakov, D. Wolf, S. R. Phillpot, A. K. Mukherjee, H. Gleiter, Nature Mater. 1, 45 (2002).
24. H. V. Swygenhoven, P. M. Derlet, A. Hasnaoui, Phys. Rev. B 66, 024101 (2002).
25. M. W. Chen et al., Science 300, 1275 (2003).
26. J. P. Hirth, J. Lothe, Theory of Dislocations (Wiley, New York, ed. 1, 1982), p. 837.
27. L. E. Murr, Interfacial Phenomena in Metals and Alloys (Addison Wesley, Reading, MA, 1975).
28. P. S. Dobson, P. J. Goodhew, R. E. Smallman, Philos. Mag. 16, 9 (1967).
29. Supported by NSF grant CMS-0140317 to the University of Pittsburgh. The work at NCEM/LBNL and at SNL was supported by the Director, Office of Science, Office of Basic Energy Sciences, Division of Materials Sciences and Engineering, of the U.S. Department of Energy. Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000. We thank L. Lu and C. Y. Song for fruitful discussion.

Pinpointing the Source of a Lunar Meteorite: Implications for the Evolution of the Moon

Edwin Gnos,[1]* Beda A. Hofmann,[2] Ali Al-Kathiri,[1]
Silvio Lorenzetti,[3] Otto Eugster,[3] Martin J. Whitehouse,[4]
Igor M. Villa,[1] A. J. Timothy Jull,[5] Jost Eikenberg,[6]
Bernhard Spettel,[7] Urs Krähenbühl,[8] Ian A. Franchi,[9]
Richard C. Greenwood[9]

The lunar meteorite Sayh al Uhaymir 169 consists of an impact melt breccia extremely enriched with potassium, rare earth elements, and phosphorus [thorium, 32.7 parts per million (ppm); uranium, 8.6 ppm; potassium oxide, 0.54 weight percent], and adherent regolith. The isotope systematics of the meteorite record four lunar impact events at 3909 ± 13 million years ago (Ma), ~2800 Ma, ~200 Ma, and <0.34 Ma, and collision with Earth sometime after 9.7 ± 1.3 thousand years ago. With these data, we can link the impact-melt breccia to Imbrium and pinpoint the source region of the meteorite to the Lalande impact crater.

The elevated Th content of the lunar Procellarum terrane was recognized during the Apollo gamma-ray remote mapping program (1). The terrane, which includes Mare Imbrium and Mare Procellarum, is characterized by K-REE-P (or KREEP) [potassium (K), rare earth elements (REE), and phosphorus (P)] rock, which is enriched in incompatible elements (2–5). Such material was returned by all Apollo missions (6). KREEP-rich material is confined to areas surrounding the Imbrium basin and the Montes Carpatus–Lalande belt (7), where the impact deposits provide a datable marker of lunar stratigraphy (6, 8). The age of the Imbrium basin has been inferred from isotope-system shock-resetting data and is debated to be either 3770 or 3850 million years old (9, 10).

Sayh al Uhaymir (SaU) 169 is a 206.45-g rock found in the Sultanate of Oman (11). The rock consists of two lithologies (Fig. 1). About 87% by volume (estimates based on

tomographic sections) (fig. S1) consists of a holocrystalline, fine-grained polymict impact-melt breccia (stage I in Fig. 1) containing 25 to 40 vol % of shocked rock and mineral clasts. The rock clasts are coarse-grained norites, gabbronorites, and mafic granulites. Crystal clasts comprise plagioclase (An_{57-94}), orthopyroxene (En_{44-78} Wo_{1-7}), olivine (Fo_{58-67}), and minor ilmenite, clinopyroxene, spinel, tridymite, or kamacite (table S1). The mineral chemistry of clasts suggests the presence of norites to olivine norites and subordinate, more evolved magmatites (granodiorites to granites) and a few granulites at the impact area. No highland anorthosites ($An_{>97}$) were identified.

The fine-crystalline impact melt (generally <200-μm grain diameter) consists of short-prismatic, low-Ca pyroxene ($En_{61-64}Wo_{3-4}$), mildly shocked plagioclase (An_{75-81}), and potassium feldspar, ilmenite, whitlockite, olivine (Fo_{58-59}), zircon, troilite, kamacite, and tridymite (see also table S1). The impact melt ilmenites contain unusually high Nb_2O_5 (~0.5 wt %), which distinguishes them from known lunar ilmenites (6). The impact melt is partially rimmed by shock-lithified regolith (13 vol%). We distinguish two stages of regolith formation (II and III in Fig. 1) from compositional differences. Regolith II and III comprise clasts of crystalline and glassy volcanic rocks, magmatic rocks, breccias, mafic granulites, and crystal fragments. Only stage III regolith is bordered by flow-banded glass and contains yellow and orange glass fragments (including glass beads), olivine basalts, pyroxferroite-bearing basalts, and an anorthosite clast. The basaltic clasts encompass the full range of compositions from Ti-rich to Ti-poor basalts, including aluminous members and picrobasalts (table S2). Glassy shock veins (stage IV in Fig. 1)

[1]Institut für Geologie, Universität Bern, Baltzerstrasse 1, CH-3012 Bern, Switzerland. [2]Naturhistorisches Museum der Burgergemeinde Bern, Bernastrasse 15, CH-3005 Bern, Switzerland. [3]Physikalisches Institut, Abteilung für Weltraumforschung und Planetologie, Universität Bern, Sidlerstrasse 5, CH-3012 Bern, Switzerland. [4]Laboratory for Isotope Geology, Swedish Museum of Natural History, Box 50007, SE-104 05, Stockholm, Sweden. [5]National Science Foundation–Arizona Accelerator Mass Spectrometry Laboratory, University of Arizona, 1118 East Fourth Street, Tucson, AZ 85721, USA. [6]Paul Scherrer Institut, 5232 Villigen, Switzerland. [7]Max-Planck-Institut für Chemie, Abteilung Kosmochemie, 55020 Mainz, Germany. [8]Departement für Chemie und Biochemie, Universität Bern, Freiestrasse 3, CH-3012 Bern, Switzerland. [9]Planetary and Space Sciences Research Institute, Open University, Milton Keynes MK7 6AA, UK.

*To whom correspondence should be addressed: gnos@geo.unibe.ch

657

crosscut impact melt and regolith and record the latest impact event.

The chemical composition of the meteor- ite has been characterized by using comple- mentary methods. Nondestructive γ-ray spec- troscopy of the uncut rock yielded a typically lunar K/U ratio of 554. The mean of three oxygen isotope measurements of the impact melt ($\delta^{17}O = 3.37$‰, $\delta^{18}O = 6.48$‰, and $\Delta^{17}O = 0.001 \pm 0.032$‰) plots on the Earth-Moon fractionation line (12). Chemical anal- yses were obtained from the impact melt, the regolith, and a large breccia clast (Table 1 and Fig. 1). Th, U, and K concentrations are similar for powder aliquots of the impact melt and for the bulk rock. A lunar origin for the meteorite is indicated by lunar element ratios (6) such as Fe/Mn (moon, 60 to 80; impact melt, 79; regolith, 78; regolith clast, 74) and K/U ratios (moon, 320 to 3800; impact melt, 535; regolith, 1682; clast, 1253).

The SaU 169 impact-melt breccia is chemically similar to Apollo KREEP impact-melt breccias (4), with slightly lower Al and Si, and higher Na, Ti, and P. However, with 32.7 ppm Th, 8.6 ppm U, and 1332 ppm REE$_{tot}$, the impact melt is more enriched in KREEP elements than is any other known lunar rock (6). Mean concentrations in Apol- lo 14 to 17 Th-rich impact-melt breccias range from 8.2 to 16.7 ppm (13). Relative to Apollo KREEP-rich impact melts, it is en- riched in incompatible trace elements by a factor of 2 to 4 (Fig. 2). The alkali elements K, Rb, and Cs occur at levels typical for Apollo impact melts but are depleted relative to other incompatible trace elements: K/Th, Rb/Th, and Cs/Th are only 0.35 ± 0.02 of the Apollo impact-melt breccia values (Fig. 2), which indicates a decoupling of the alkalis during incompatible trace-element fraction- ation. A low Rb/Sr of 0.04 shows that it is not related to highly evolved KREEP differenti- ates (Rb/Sr, 0.10 to 0.56) (13). The sidero- phile elements Ni, Co, Ir, and Au are present at levels similar to Apollo impact melts and require ~0.5% of a meteoritic component.

The average regolith (Fig. 1 and Table 1) has a composition similar to Apollo 12, 14,

Table 1. Whole-rock chemistry.

Oxides or elements	Impact-melt breccia 827 mg*	Average regolith 185 mg†	KREEP clast 117 mg†
	wt %		
SiO$_2$	45.15	46.9‡	–
Al$_2$O$_3$	15.88	17.54	16.34
FeO$_{tot}$	10.67	11.09	8.80
MnO	0.14	0.14	0.12
MgO	11.09	7.94	6.92
CaO	10.16	11.72	10.60
Na$_2$O	0.98	0.78	1.18
K$_2$O	0.54	0.46	0.88
TiO$_2$	2.21	2.49	1.47
ZrO$_2$	0.38	0.08	0.19
BaO	0.17	0.07	0.15
Cr$_2$O$_3$	0.14	0.19	0.12
P$_2$O$_5$	1.14	0.42	0.76
S	0.33		
Total	98.98	99.82	
	ppm		
Sc	25	28	18
V	36	61	36
Cr	992	1310	811
Co	31	19	12
Ni	204	82	58
Cu	9	<20	<20
Zn	31	<60	<60
Rb	13.7	10.0	20.0
Sr	359	214	230
Y	532	162.5	338
Zr	2835	596	1397
Nb	124	18	112
Cs	0.8	0.4	0.9
Ba	1520	593	1351
La	170	52	113
Ce	427	139	297
Pr	57.45	17.1	35.6
Nd	256.5	76.9	162
Sm	70.15	21.9	44.9
Eu	4.20	2.43	2.45
Gd	86.4	25.3	50.4
Tb	15.1	5.08	10.5
Dy	94.15	30.7	63.9
Ho	21.3	6.36	13
Er	58.05	18.6	39.3
Tm	9.13	2.72	5.96
Yb	54.65	16.9	36
Lu	7.64	2.53	5.24
Hf	64.3	14.8	34.7
Ta	7.1	2.14	4.16
W	3.45	1.3	2.5
Ir ppb	4.2		
Au ppb	6		
Pb	13.8	<10	<10
Th	32.7	8.44	21.70
U	8.6	2.27	5.83
Sum REE (ppm)	1332	418	879
Fe/Mn (wt)	76.5	79.5	73.6
Mg/(Fe+Mg) (mol)	0.65	0.56	0.58
K/U (wt)	521	1682	1253
Th/U (wt)	3.80	3.72	3.72

*Combined inductively coupled plasma mass spectrometry and optical emission spectrometry, and instrumental neu- tron activation analysis. †Inductively coupled plasma mass spectrometry and optical emission spectrometry. ‡Microprobe value on regolith shock glass.

Fig. 1. Thin-section photograph showing the different lithologies and their age relationship: I, impact-melt breccia. II and III, two types of regolith (III intrudes II). IV designates shock veins as youngest formation. Ol, olivine; Plg, plagioclase; Opx, orthopyroxene. Image width is 35 mm.
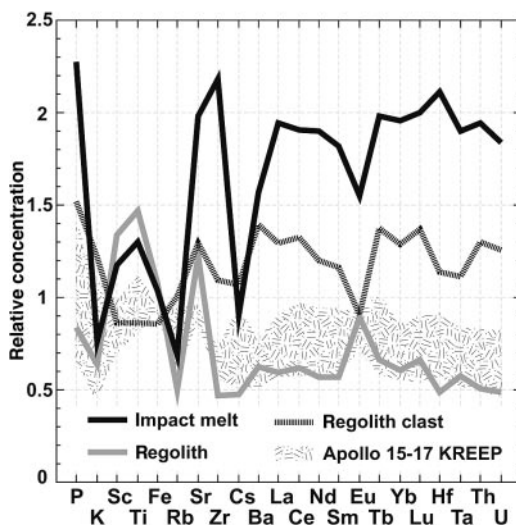
Fig. 2. Element abundances of impact-melt breccia, regolith II and III combined, and a polymict breccia clast in regolith II normalized to average Apollo 14 high-Th impact-melt brec- cias (13). Data for Apollo 15 to 17 impact melts are shown for compari- son. Compared with Apollo data, P, Zr, REE, Th, and U contents are signifi- cantly higher, but K, Cs, and Rb, as well as Al and Si (typically fractionat- ed in potassium feldspar), are lower.

and 15 regoliths. In contrast to the impact melt, the regolith is not depleted in Na and K (Fig. 2), as shown by the higher K/U of 1682. The regolith clast is an evolved KREEP rock (Fig. 2) with 21.7 ppm Th, but it is unrelated to the impact melt, on the basis of its K/U ratio of 1253. The exceptionally high Th, U, and K contents restrict the provenance of SaU 169 to the Imbrium area (*13*).

We determined the ages of the four impact events using isotopic data. ^{207}Pb-^{206}Pb isotope ratios (table S3) were obtained in thin section by ion-microprobe analysis on 12 poikilitic impact-melt zircons and yielded a crystallization age of 3909 ± 13 Ma [2σ; mean square weighted deviation (MSWD), 0.33] (Fig. 3). ^{39}Ar-^{40}Ar data (table S4) were obtained on a feldspar concentrate from the impact melt. The irregular age spectrum in combination with the Ca/K concentrations indicate a resetting of the Ca-rich feldspars at ~2800 Ma and a younger disturbance at <500 Ma affecting the potassium feldspars (fig. S2).

The duration of cosmic-ray exposure (CRE) of the impact-melt breccia and the younger regolith (stage III in Fig. 1) was obtained from light noble gas isotopic characteristics (table S5). All three samples analyzed show He loss; hence, the ^{3}He CRE age is not significant. The production rates for ^{21}Ne and ^{38}Ar depend on the chemical composition and on shielding depth during exposure to cosmic rays. The $(^{22}$Ne/^{21}Ne$)_c$ value of 1.197 for the impact melt and 1.200 for the regolith indicate irradiation at a shielding depth of a few tens of centimeters. Using the method for calculating production rates proposed by (*14*) at the lunar surface and (*15*) during transfer in space, and assuming a typical shielding of 40 g/cm², we obtained lunar surface CRE ages from ^{21}Ne and ^{38}Ar of 200 ± 40 Ma and 182 ± 36 Ma for the impact melt, and 150 ± 30 Ma and 192 ± 38 Ma for the regolith. The ^{38}Ar CRE age extracted from the ^{39}Ar-^{40}Ar data is also 192 ± 20 Ma (table S4 and fig. S2).

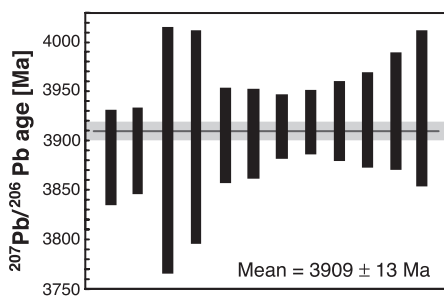By adopting a saturation activity for ^{10}Be of 25 dpm/kg as given by (*16*), we found the

upper time limit for a small rock exposure during Moon-Earth transfer, estimated from ^{10}Be (table S6), to be 0.34 Ma. Thus, the exposure in free space is negligibly short compared with the total CRE. We conclude that the CRE ages are lunar regolith residence times. Like most lunar meteorites (*17*), SaU 169 thus seems to have resided in the upper meter of the regolith before launch into space, as supported by our CRE ages and by the solar wind component found in the stage III regolith (table S5). The terrestrial age obtained by ^{14}C and ^{10}Be methods is ≤9700 ± 1300 years (table S6).

The combined age data indicate a complex lunar history. Crystallization of the impact melt occurred at 3909 ± 13 Ma, followed by exhumation by a second impact at ~2800 Ma, which raised the sample to a regolith position at unconstrained depth. A third impact at ~200 Ma moved the material closer to the lunar surface, where it mixed with solar-wind–containing regolith. It was launched into space by a fourth impact at <0.34 Ma.

The 3909 ± 13 Ma ^{207}Pb-^{206}Pb zircon age on the impact melt is an accurate estimate for the Imbrium impact event. The small difference from the 3770 ± 20 or 3850 ± 20 Ma range obtained by whole-rock ^{39}Ar-^{40}Ar system (*9*, *10*, *18*) confirms that the argon system yields reliable age estimates. The age difference may reflect the imprecisely known ^{40}K decay constant. A recalculation of the argon ages with the decay value proposed by (*19*) results in a ~1% age increase.

We were able to constrain the SaU 169 provenance even more tightly by using Lunar Orbiter images in combination with Clementine-derived Fe-Ti maps, Lunar Prospector Th maps, Clementine color-ratio images (750/950 nm), lunar crater ages (*20*, *21*) (table S7), and data from the Apollo and Luna missions. The source location is based on the following assumptions: (i) The high Th (32.7 ppm) impact melt is derived from one of the highest Th areas defined by Lunar Prospector γ-ray mapping; (ii) The regolith Fe-Ti-Th concentrations are representative of the regolith of the ejection area; (iii) The age disturbances at ~2800 and 200 Ma were caused by impacts that were able to disturb the Ar system; (iv) The ejection crater from which the rock was launched to Earth is on the order of a few kilometers (*17*).

The Lunar Prospector γ-ray mapping demonstrated that the surface occurrence of KREEP is confined to the near-side Procellarum terrane (*2–5*, *22*, *23*). Because the bulk Th content of the impact melt (Table 1) is higher than any pixel of the Lunar Prospector measurements (maximum, 9.36 to 10.40 ppm) (*7*), we assume that SaU 169 must come from one of the Th hot spots (fig. S3), which are located at the Aristarchus and Aristillus

craters, the Montes Carpathus–Fra Mauro region, and the Lalande crater region (*7*). Among these Th hot spots, only the area around Lalande and southeast of the crater Aristillus are compatible with our regolith Fe-Ti-Th concentrations (table S7). In both areas, a young (bright-halo), small crater as a possible launch-to-Earth crater is evident. However, the age data for SaU 169 render the Lalande crater area (9°W 5°S) the most plausible because it contains a ~2800 Ma crater (Lalande, 2246 to 2803 Ma) and a ~200 Ma crater (e.g., Lalande A, 175 to 300 Ma) (table S7). Based on the comprehensive data set presented, we conclude that an origin from the Lalande crater area is most likely.

SaU 169 is the only large high-KREEP sample for which simultaneous determinations of precise age and chemistry are available. The Imbrium impact age dates a strong decline of large impacts in the inner solar system corresponding to the beginning of life-supporting conditions on Earth.

References and Notes

1. A. E. Metzger, J. I. Trombka, L. E. Peterson, R. C. Reedy, J. R. Arnold, *Science* **179**, 800 (1973).
2. L. A. Haskin, J. L. Gillis, R. L. Korotev, B. L. Jolliff, *J. Geophys. Res.* **105**, 20403 (2000).
3. B. L. Jolliff, J. J. Gillis, L. A. Haskin, R. L. Korotev, M. A. Wieczorek, *J. Geophys. Res.* **105**, 4197 (2000).
4. R. L. Korotev, *J. Geophys. Res.* **105**, 4317 (2000).
5. M. A. Wieczorek, R. J. Phillips, *J. Geophys. Res.* **105**, 20417 (2000).
6. G. H. Heiken, D. T. Vaniman, B. M. French, *Lunar Source Book: A User's Guide to the Moon* (Cambridge University Press, Cambridge, 1991).
7. D. J. Lawrence *et al.*, *Geophys. Res. Lett.* **26**, 2681 (1999).
8. D. E. Wilhelms, *U.S. Geol. Survey Prof. Pap.* **1348**, 1 (1987).
9. B. A. Cohen, T. D. Swindle, D. A. Kring, *Science* **290**, 1754 (2000).
10. D. Stöffler, G. Ryder, *Space Sci. Rev.* **96**, 9 (2001).
11. S. Russell *et al.*, *Meteorit. Planet. Sci.* **38**, A189 (2003).
12. U. Wiechert *et al.*, *Science* **294**, 345 (2001).
13. B. L. Jolliff, *Int. Geol. Rev.* **40**, 916 (1998).
14. C. Hohenberg, K. Marti, F. Podosek, R. Reedy, J. Shirck, *Proc. Lun. Planet. Sci. Conf. 9th*, 2311 (1978).
15. O. Eugster, T. Michel, *Geochim. Cosmochim. Acta* **59**, 177 (1995).
16. C. Tuniz *et al.*, *Geophys. Res. Lett.* **10**, 804 (1983).
17. P. H. Warren, *Icarus* **111**, 338 (1994).
18. G. B. Dalrymple, G. Ryder, *J. Geophys. Res.* **101**, 26069 (1996).
19. K. Min, R. Mundil, P. R. Renne, K. R. Ludwig, *Geochim. Cosmochim. Acta* **64**, 73 (2000).
20. R. B. Baldwin, *Icarus* **71**, 19 (1987).
21. R. B. Baldwin, *Icarus* **61**, 63 (1999).
22. R. C. Elphic *et al.*, *J. Geophys. Res.* **107**, **E4**, 8-1 (2002).
23. W. C. Feldman, O. Gasnault, S. Maurice, D. J. Lawrence, R. C. Elphic, *J. Geophys. Res.* **107**, **E3**, 5-1 (2002).
24. Materials and methods are available as supporting material on *Science* Online.
25. Funded by Swiss National Foundation grants 21-64929.01, 20-61933.00, and credit 21-26579.89. Thanks to H. Al-Azri and K. Musallam for support; to L. R. Gaddis, USGS, for Lunar Orbiter images; and to J. Kramers, J. Ridley, L. Diamond, D. Fleitmann, T. Armbruster, and B. Hacker for reviews.

Fig. 3. ^{207}Pb/^{206}Pb ages of 12 zircons from the impact melt. Data are corrected for common Pb (*24*) and yield a weighted average age of 3909 ± 13 Ma (2σ; MSWD = 0.33).

Foundering Lithosphere Imaged Beneath the Southern Sierra Nevada, California, USA

Oliver S. Boyd,* Craig H. Jones, Anne F. Sheehan

Seismic tomography reveals garnet-rich crust and mantle lithosphere descending into the upper mantle beneath the southeastern Sierra Nevada. The descending lithosphere consists of two layers: an iron-rich eclogite above a magnesium-rich garnet peridotite. These results place descending eclogite above and east of high P wave speed material previously imaged beneath the southern Great Valley, suggesting a previously unsuspected coherence in the lithospheric removal process.

The Sierra Nevada mountain range of California is one of the highest (about 3 km mean elevation) in the United States; however, the crust is only 35 km thick (1) and requires some unusual structure in the mantle. Xenoliths (2) and volcanic rocks (3, 4) suggest that the upper mantle source for these materials, beneath the Sierra Nevada, changed from an eclogite facies garnet pyroxenite to a garnet-free spinel peridotite about 4 million years ago (Ma). Eclogitic rocks may still be present in the upper mantle below the Sierra Nevada (5), and knowing where and how much is left constrains how lithospheric material is removed and how continental crust grows and changes beneath mountain ranges (6). To determine whether there are any eclogitic rocks left, we used three-dimensional models of P and S wave speeds to define compositional and thermal characteristics of the lithosphere. We measured attenuation to correct the velocities for anelastic effects, which include temperature and hydration. The corrected wave speeds are subsequently termed anharmonic wave speeds and compared with laboratory predictions.

Low Pn velocities directly beneath the high Sierra (7, 8) and high attenuation in the upper mantle in the region (9) are consistent with high temperature, low-density material. Shear wave splitting measurements determined from the seismic phase SKS (10, 11) (Fig. 1) indicate strong seismic anisotropy oriented N80°E beneath the central Sierra, probably because of the presence of strained, olivine-rich, peridotitic mantle. Less splitting beneath the eastern and western Sierra may indicate largely isotropic material, such as eclogite (12), little strain, or a vertical fast axis of anisotropy. The region of large SKS splitting and low Pn velocity correlates well with high mantle electrical conduc-

tivity that is probably due to 1% partial melt beneath the southern Sierra Nevada (13). These observations are consistent with the presence of a peridotitic uppermost mantle near asthenospheric temperatures near the Sierran crest today.

We recorded 40 teleseismic events that yielded 800 seismic wave traces on 24 broadband seismometers (Fig. 1). We measured

direct P wave and S_{fast} and S_{slow} wave travel time delays (t_P, t_{S_f}, and t_{S_s}) and S_{fast} and S_{slow} wave path–integrated attenuation, Δt^* (14), to invert for variations in P wave speed, v_P, fast and slow S wave speeds (v_{S_f} and v_{S_s}), and fast- and slow-oriented S wave attenuation ($Q_{S_f}^{-1}$ and $Q_{S_s}^{-1}$). These were combined to examine variations in the ratio of P wave to S wave velocity, v_P/v_S, and transverse S wave anisotropy. Our tomographic inversions reduce the variance of the data by more than half of that available, whereas measurement noise and station and event statics account for the remainder (15).

The tomographic models reveal regions of alternating high and low velocity and attenuation that dip ~55° to the east southeast (Fig. 2). The region of high P wave velocity at 36.4°N extends from –119.5°W to –118°W and may be what has been referred to as the Isabella anomaly. The volume above this anomaly consists of a slow P and S wave region. The S wave velocities decrease more across this contrast than do the P wave velocities, giving rise to an increase in the v_P/v_S ratio. These seismic anomalies trend northeast (Fig. 3C).
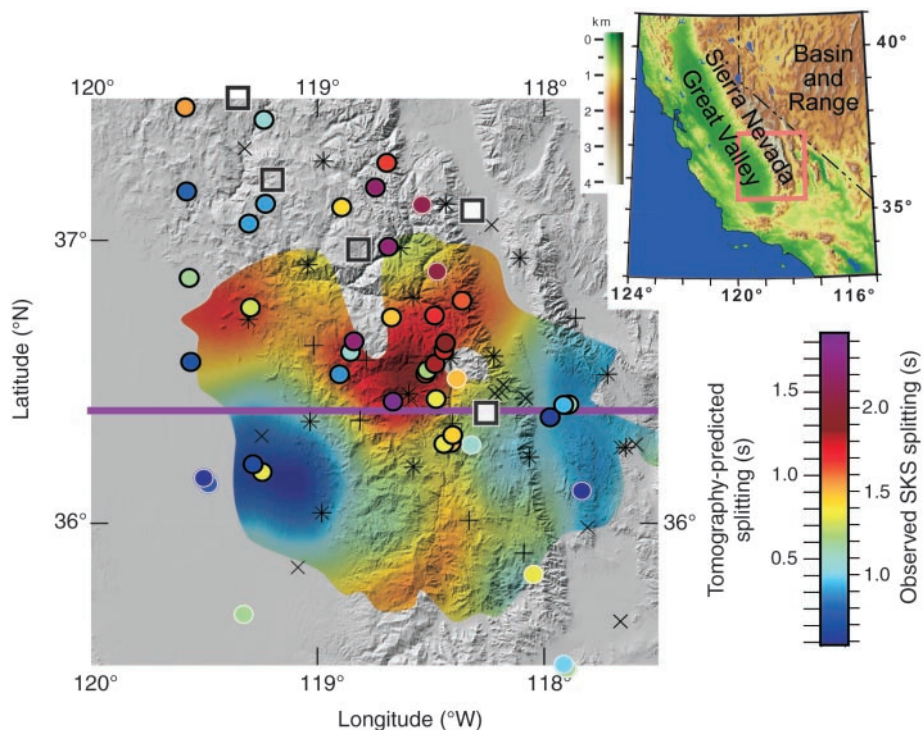

Fig. 1. Topographic map of the southern Sierra Nevada overlain with tomographically predicted shear wave splitting (this study) and independently determined SKS splitting measurements, filled circles with white (10) or black (11) outline. Map of Nevada and California with pink outline of study area given in the upper right for reference. The SKS measurements are placed at 125-km depth along the SKS ray path. The tomographically predicted splitting time is determined by integrating differences between the fast and slow S wave speed models in the upper 200 km for regions with greater than 150-km thickness of resolution greater than 0.5. Thick purple line crossing at 36.4°N is the latitudinal position of the vertical slices in Fig. 2. The plus symbols are seismograph stations used for the tomography, the multiplication symbols are stations used to measure SKS splitting, and the asterisks, plus symbols overlain by multiplication symbols, are stations used for both tomography and SKS splitting. The open boxes are garnet peridotite and garnet pyroxenite xenolith localities; solid boxes are spinel peridotite xenolith localities.

Department of Geological Sciences, University of Colorado at Boulder, 2200 Colorado Avenue, Boulder, CO 80309–0399, USA.

*To whom correspondence should be addressed. E-mail: oliver.boyd@colorado.edu

Upon vertically integrating the differences in S_{fast} and S_{slow} travel times for the top 200 km, we are able explain more than 50% of the variance of the SKS splitting measurements (Fig. 1 and fig. S1) (15), an independent verification of our subsurface distribution in anisotropy. The SKS measurements were not used in the inversion; we used only the orientation of fast S wave speed derived from the splitting analysis. The shallow depth for the origin of the splitting (<200 km) is consistent with the 1-s variation in splitting amplitudes over short distances (<50 km) and inconsistent with splitting distributed over greater depths.

Attenuation is needed to distinguish between the effects on wave speed from thermal and compositional variations (16, 17). Anisotropy (12) and Poisson's ratio (18) can then be used to constrain the composition (Table 1). Because of anelasticity, temperature variations will cause seismic velocities and attenuation to vary predictably (19). We interpret deviations from the thermal relationship as being caused by variations in composition.

Because the dipping layer of low wave speeds is accompanied by low attenuation, temperatures are relatively low (~200 K lower than the material to the east), and partial melt or partially water-saturated minerals are not likely present. This leaves some sort of compositional effect producing, in addition to low velocities, a high v_P/v_S ratio and variable anisotropy. High v_P/v_S ratios can indicate an increase in the amount of pyroxene and garnet relative to olivine or a decrease in the Mg number [Mg/(Mg + Fe)]. The variable anisotropy is more difficult to interpret. Lower

anisotropies can be from an absence of olivine, an absence of strain, or a heterogeneously strained peridotite. In response to infilling asthenosphere, the low anisotropy in the inferred region of spinel peridotite (Fig. 2D) could be anisotropic but have a near-vertical axis of orientation. Normally isotropic lithologies like eclogites might be anisotropic if pervasively cut by dikes or if the pyroxenes have become preferentially aligned (15). The large anisotropy appearing below the dipping layer of high P wave velocities may reflect lateral extrusion of asthenospheric peridotite in response to delaminating mantle lithosphere.

Ambiguity in interpreting the tomographic results can be reduced by examining the xenoliths erupted nearby in the central Sierra Nevada. There are three groups of xenoliths: garnet pyroxenites and garnet peridotites (both erupted before 8 Ma) constituting the old mantle lithosphere (2, 20) and fertile, olivine-rich, garnet-free spinel peridotites (erupted since 4 Ma) with an asthenospheric affinity (2). Chemically, the garnet and pyroxene in the pyroxenites tend to have low Mg, whereas the olivine in the peridotites has a higher Mg number (2). This observation means that the garnet pyroxenite will have higher v_P/v_S ratios relative to the peridotites.

We calculated the P and S wave speeds (15, 21) for a range of expected compositions and compare these values to our observations of anharmonic wave speed (15) to broadly determine the mineralogy and Mg numbers in the observed seismic anomalies (Fig. 3). The modal proportions for the garnet pyroxenite xenolith samples are predicted to be about 65% clinopyroxene and 35% garnet (22). The Mg number ranges from 0.6 to 0.9 for the clinopyroxenes and 0.3 to 0.6 for the garnet (2). The composition that best matches our seismic observations averages 70% pyroxene and 30% garnet (15). Compositions having as much as 40% garnet would be acceptable if a Reuss average, as opposed to the Voight-Reuss-Hill average (23) used here, is favored. One garnet peridotite sample has 75% olivine, 15% orthopyroxene, 5% clinopyroxene, and 5% garnet (24), whereas the Mg numbers are reported to be on average 0.91, 0.91, 0.92, and 0.85 for those minerals, respectively (2). Our seismically determined composition averages 75% olivine, 20% pyroxene, and 5% garnet. The young peridotite samples are devoid of garnet and average 80% olivine, with the remainder orthopyroxene, clinopyroxene, and spinel (25). Their Mg numbers average 0.89, 0.89, 0.87, and 0.67, respectively (2). Because a majority of the young samples have been classified as harzburgites, we restrict our analysis to clinopyroxene-free assemblages. We find this composition has on average 80% olivine, 16% pyroxene, and 4% garnet (26).
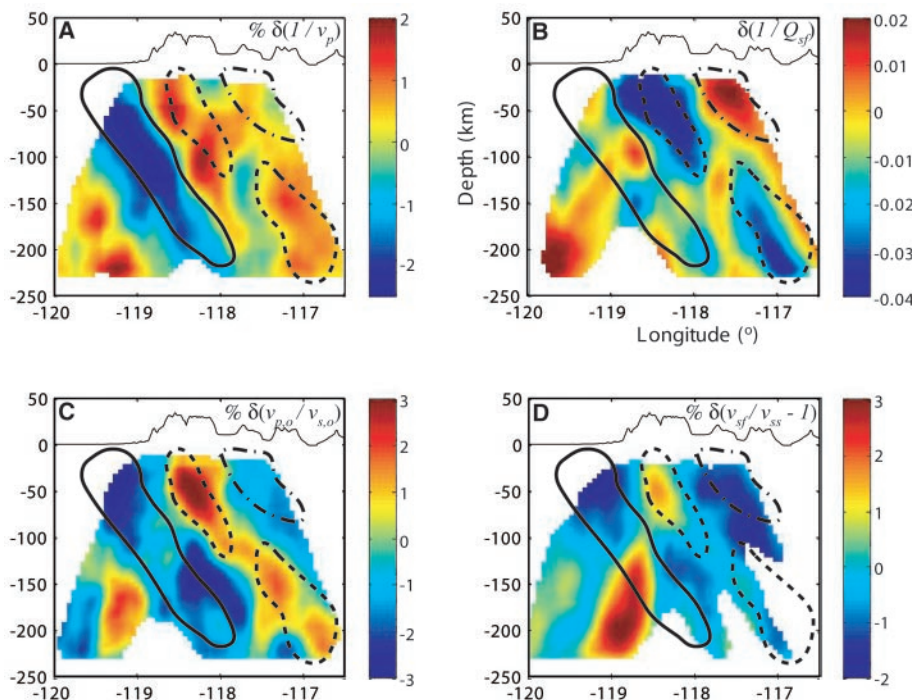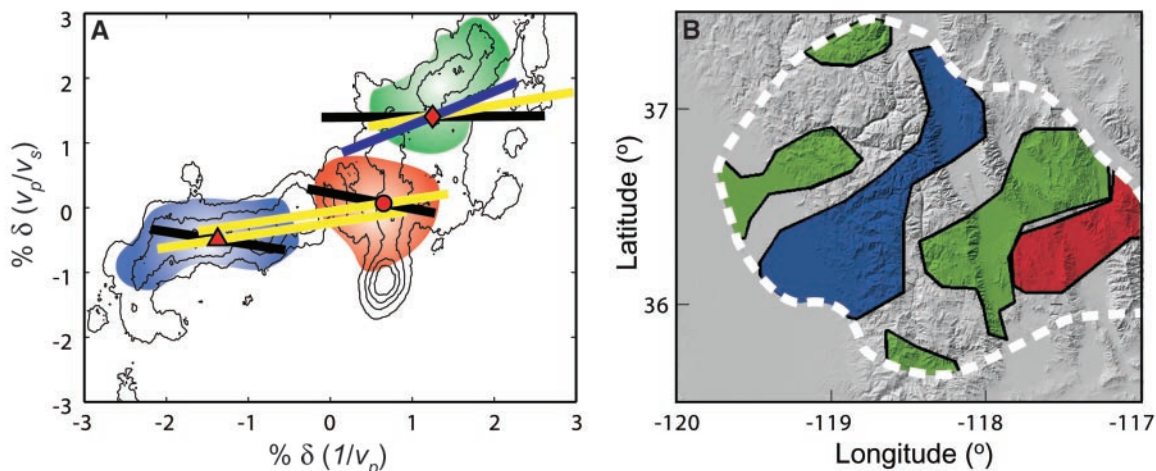

Fig. 2. Vertical slices of our tomographic models and derived quantities: percent change in P wave slowness (**A**), change in attenuation (**B**), percent change in anharmonic v_P/v_S ratio where v_S is taken from the average of the fast and slow models (**C**), and shear wave anisotropy where the slow model is derived as residuals from the fast model (**D**). The solid lines indicate regions of descending garnet peridotite, and the dashed lines delineate regions of garnet pyroxenite. The regions of low velocities and high attenuation above the garnet pyroxenite are presumably the infilling of asthenospheric spinel peridotite (dashed-dot outlines). Topography is depicted with 10 times vertical exaggeration.

Table 1. Changes observed in geophysical properties in the mantle from different causes. The asterisk indicates that the attenuation will only increase if high temperatures are required to produce melt (15). The question mark indicates that anisotropy may increase or change orientation if melt pockets and/or bands are favorably oriented.

Increasing factor	v_P	v_P/v_S	Attenuation	Anisotropy	Density
Temperature	Decrease	Increase	Increase	No change	Decrease
Melt	Decrease	Increase	*	?	Decrease
Magnesium	Increase	Decrease	No change	No change	Decrease
Garnet/olivine ratio	Increase	Increase	No change	Decrease	Increase

Fig. 3. (**A**) Percent change in anharmonic v_P/v_S ratio versus percent change in anharmonic P wave slowness for a range of model compositions at a pressure of 3 GPa and a range of temperatures. Compositions defined in the text that best match the seismic observations are garnet pyroxenite (red diamond) at 1200°C, spinel peridotite (red circle) at 2000°C, and garnet peridotite (red triangle) at 1000°C. Thick black lines represent ranges in composition where garnet is substituted for pyroxene. For the garnet pyroxenite, garnet varies from 20% (right end of black line, slower wave speeds) to 40%; the spinel peridotite, 0 to 10%; and the garnet peridotite, 0 to 10%. Yellow lines represent a 400 K variation in temperature (900 to 1300°C), with faster wave speeds (to the left) accompanying lower temperatures. The blue line spans the range of Mg numbers for the garnet pyroxenite samples. Contours denote anharmonic P wave slowness and v_P/v_S from our Sierran tomography from 75- to 100-km depth. Each contour interval represents 20% of the data. Colored areas correspond to compositional regions outlined in the tomography of Fig. 2 and shaded areas in the 100-km-depth slice (**B**). Garnet pyroxenite, green; spinel peridotite, red; and garnet peridotite, blue. Compositional zones in (B) derived from tomography using composition relations (*15*) with P wave slowness, Poisson's ratio, anisotropy, and attenuation. The white dashed line delimits the part of the tomography model with resolution greater than 0.4.

Many eclogites are seismically fast because they have a large proportion of garnet and much of the pyroxene is in the form of jadeite (*27*). The xenoliths found here, however, have large amounts of clinopyroxene (*24*) and very little jadeite (*2*). The garnet pyroxenites are therefore slower than common eclogites and have a higher v_P/v_S ratio relative to the peridotites (Fig. 3A). A comparison of the characteristics of the seismic anomalies to calculated values for mineral assemblages (Fig. 3A) reveals that the dipping layer of high velocity is the garnet peridotite; the low velocity layer above, garnet pyroxenite; and the seismically intermediate region above and to the east, asthenospheric spinel peridotite (Fig. 2). A map view of the older garnet-bearing assemblages at 100-km depth indicates a northeast strike to this structure (Fig. 3B). The seismically determined compositions imply that, at 100-km depth, densities increase from close to neutrally buoyant garnet peridotite below to very dense, negatively buoyant ($\Delta\rho \approx 185$ kg/m^3) garnet pyroxenite above.

Although we have suggested specific differences in composition and temperature for the three rock assemblages identified, the most robust conclusion is that three distinct rock types exist in the upper mantle beneath the Sierra Nevada. This observation is made possible by correcting measured P and S wave seismic velocities for attenuation and comparing these values to laboratory results. The apparent dip of the two garnet-bearing packages and the asymmetry of the package (garnet peridotite to the west of the garnet pyroxenite) are important constraints on the mechanism of removal of this material. The asymmetry resembles delamination of stratified mantle lithosphere from the crust and is inconsistent with most numerical models of removal by development of a Rayleigh-Taylor instability (*28*), although such asymmetry might be produced by an initially asymmetric instability (*29*). The eastward plunge of this body indicates that either material has moved directly downward or somewhat to the east relative to the overlying crust, in contrast to motion to the west relative to the crust inferred from older images of this structure (*5*). Imaging one of the very few (if not the only) examples of ongoing removal of mantle lithosphere from beneath continental crust, this tomography provides some of the best observational constraints on the process of removing dense material from continental crust.

References and Notes

1. B. Wernicke *et al.*, *Science* **271**, 190 (1996).
2. M. Ducea, J. Saleeby, *J. Geophys. Res.* **101**, 8229 (1996).
3. C. R. Manley, A. F. Glazner, G. L. Farmer, *Geology* **28**, 811 (2000).
4. G. L. Farmer, A. F. Glazner, C. R. Manley, *Geol. Soc. Am. Bull.* **114**, 754 (2002).
5. G. Zandt, *Int. Geol. Rev.* **45**, 213 (2003).
6. M. Ducea, *J. Geophys. Res.* **107**, 10.1029/2001JB000643 (2002).
7. C. H. Jones, H. Kanamori, S. W. Roecker, *J. Geophys. Res.* **99**, 4567 (1994).
8. M. M. Fliedner *et al.*, *Geology* **24**, 367 (1996).
9. H. H. Al-Khatib, B. J. Mitchell, *J. Geophys. Res.* **96**, 18129 (1991).
10. J. Polet, H. Kanamori, *Geophys. J. Int.* **149**, 313 (2002).
11. C. H. Jones, R. A. Phinney, *Geol. Soc. Am. Abstr. Prog.* **31** (no. 7), A-481 (1999).
12. N. I. Christensen, W. D. Mooney, *J. Geophys. Res.* **100**, 9761 (1995).
13. M. Ducea, S. K. Park, *Geophys. Res. Lett.* **27**, 2405 (2000).
14. A. F. Sheehan, S. C. Solomon, *J. Geophys. Res.* **97**, 15339 (1992).
15. Please refer to the supporting online material on *Science* Online for more information about seismic methods, resolution, and calculation of mineral properties.
16. O. S. Boyd, A. F. Sheehan, in *Lithospheric Structure and Evolution of the Rocky Mountain Region*, K. E. Karlstrom, G. R. Keller, Eds. (*Geophys. Monogr. Am. Geophys. Union*), in press.
17. S. Karato, *Geophys. Res. Lett.* **20**, 1623 (1993).
18. N. I. Christensen, *J. Geophys. Res.* **101**, 3139 (1996).
19. I. Jackson, *Annu. Rev. Earth Planet. Sci.* **21**, 375 (1993).
20. F. C. W. Dodge, J. P. Lockwood, L. C. Calk, *Geol. Soc. Am. Bull.* **100**, 938 (1988).
21. B. R. Hacker, G. A. Abers, *Geochem. Geophys. Geosys.* **5**, 10.1029/2003GC000614 (2004).
22. J. Saleeby, M. Ducea, D. Clemens-Knott, *Tectonics* **22**, 10.1029/2002TC001374 (2003).
23. J. P. Watt, G. F. Davies, J. O'Connell, *Rev. Geophys. Space Phys.* **14**, 541 (1976).
24. M. Ducea, J. Saleeby, *Contrib. Mineral. Petrol.* **133**, 169 (1998).
25. H. G. Wilshire *et al.*, *U.S. Geol. Surv. Prof. Pap. 1443* (1988), p. 179.
26. Garnet is used for the spinel peridotite composition because we are considering seismic velocities at 3 GPa, by which pressure spinel would have transformed to garnet.
27. R. G. Coleman, D. E. Lee, J. B. Beatty, W. W. Brannock, *Geol. Soc. Am. Bull.* **76**, 483 (1965).
28. G. A. Houseman, P. Molnar, *Geophys. J. Int.* **128**, 125 (1997).
29. M. Jull, P. B. Kelemen, *J. Geophys. Res.* **106**, 6423 (2001).
30. Thanks to P. Molnar and L. Farmer for helpful discussion and B. Phinney and several land use agencies (U.S. Forest Service, Sequoia and Kings Canyon National Parks, and Bureau of Land Management) for help in obtaining the data. We also acknowledge two anonymous reviewers and L. Rowan for thorough reviews. This work was supported by NSF grants 9526974 and 0003747 and a Cooperative Institute for Research in Environmental Research Sciences fellowship.

Herbivores Promote Habitat Specialization by Trees in Amazonian Forests

Paul V. A. Fine,[1,2]* Italo Mesones,[3] Phyllis D. Coley[1]

In an edaphically heterogeneous area in the Peruvian Amazon, clay soils and nutrient-poor white sands each harbor distinctive plant communities. To determine whether a trade-off between growth and antiherbivore defense enforces habitat specialization on these two soil types, we conducted a reciprocal transplant study of seedlings of 20 species from six genera of phylogenetically independent pairs of edaphic specialist trees and manipulated the presence of herbivores. Clay specialist species grew significantly faster than white-sand specialists in both soil types when protected from herbivores. However, when unprotected, white-sand specialists dominated in white-sand forests and clay specialists dominated in clay forests. Therefore, habitat specialization in this system results from an interaction of herbivore pressure with soil type.

The floras of tropical forests in areas of high environmental heterogeneity have high beta diversity, leading many authors to conclude that habitat specialization in tropical plants contributes significantly to the global diversity gradient (1–4). The change in species composition between habitats (beta diversity) in plants is thought to be caused primarily by physiological adaptations to different abiotic conditions (5–7). But herbivores have been shown to affect plant population dynamics in many ways, including plant distributions (8–10). The best way to test the effect of herbivores on plant distributions is with transplant experiments involving multiple species, but few such studies have been conducted. Although they have evaluated only one or two species at a time, these studies have consistently shown that herbivores can significantly limit plant distributions, often restricting species to a subset of the habitats they could physiologically tolerate in the absence of herbivores (11–15). However, no such studies have been conducted in tropical forests [but see (16, 17)], where herbivore abundance and rates of herbivory far exceed those of the temperate zone (18). If herbivores interact with abiotic gradients to intensify differences in habitat quality, this phenomenon could account for the high degree of habitat specialization in environmentally heterogeneous areas within tropical systems.

The lowland Amazonian ever-wet rainforest in the Allpahuayo-Mishana Reserve near Iquitos, Peru (3°57'S, 73°26'W), provides an ideal system to study habitat specialization and the role of insect herbivores. Forests in the Iquitos area occur on a mosaic of soil types with well-defined boundaries, including extremely infertile white-sand soils adjacent to lateritic red-clay soils (19). These two soil types are each characterized by a highly distinctive but related flora (19–22). Many white-sand specialist trees belong to the same genera as neighboring clay forest specialists, allowing for a phylogenetically controlled experiment using edaphic specialist species.

Janzen (23) argued that the most critical factors in the evolution of white-sand specialist plants were herbivores and plant defenses rather than physiological tolerance to poor soils. A mechanism that would produce such a result is a trade-off between allocations to growth and defense that selects for different traits in white-sand and clay forest species. We hypothesized that plants that do not invest sufficiently in defenses ought to be excluded from white-sand forests by herbivores, because the cost of replacing damaged tissue in such a nutrient-poor environment should be prohibitively high (24). In contrast, species that invest heavily in defenses should grow more slowly and be at a competitive disadvantage in clay forests.

[1]Department of Biology, University of Utah, 257 S. 1400 East, Salt Lake City, UT 84112, USA. [2]Environmental and Conservation Programs and Department of Botany, Field Museum of Natural History, 1400 S. Lake Shore Drive, Chicago, IL 60605, USA. [3]Department of Forestry, Universidad Nacional de la Amazonía Peruana, Plaza Serafín Filomeno 246, Iquitos, Peru.

*To whom correspondence should be addressed. E-mail: fine@biology.utah.edu

Fig. 1. Phylogenetic relationships of the species in the transplant experiment with their family, genus, species, and soil type (origin). The phylogenetic tree is adapted from (37), and the Burseraceae phylogeny is adapted from (38). *Tetragastris panamensis* has been found to be imbedded within *Protium* sensu lato (38), therefore we are considering *Tetragastris panamensis* as a *Protium* clay specialist species. Soil type (origin) was determined by extensive tree and seedling inventories that we conducted in more than 25 sites throughout the 57,600-ha Allpahuayo-Mishana Reserve. Our designation of white-sand specialists is consistent with other inventories and a published flora from the area (19, 22, 38, 39). Our designation of clay specialists conforms to other species lists on clay soils in the area (19, 22, 38) as well as with lists from the Ecuadorian Amazon, which is almost entirely composed of clay soils and where white-sand forests are unknown (40).

Family	Genus and species	Soil type (origin)
Annonaceae	*Oxandra xylopioides* Diels	Clay
Annonaceae	*Oxandra euneura* Diels	White sand
Burseraceae	*Tetragastris* panamensis* (Engler) Kuntze	Clay
Burseraceae	*Protium subserratum* (Engler) Engler	White sand
Burseraceae	*Protium nodulosum* Swart	Clay
Burseraceae	*Protium paniculatum* Engler	White sand
Burseraceae	*Protium opacum* Swart	Clay
Burseraceae	*Protium calanense* Cuatrec.	White sand
Burseraceae	*Protium trifoliolatum* Engler	Clay
Burseraceae	*Protium krukoffii* Swart	Clay
Burseraceae	*Protium heptaphyllum* subsp. *ulei* (Swart) Daly	White sand
Burseraceae	*Protium hebetatum* Daly	Clay
Malvaceae (Bombacoid)	*Pachira insignis* (Swartz) Swartz ex Savigny	Clay
Malvaceae (Bombacoid)	*Pachira brevipes* (Robyns) Alverson	White sand
Euphorbiaceae	*Mabea pulcherrima* Müll. Arg.	Clay
Euphorbiaceae	*Mabea subsessilis* Pax & Hoffmann	White sand
Fabaceae (Faboid)	*Swartzia arborescens* (Aublet) Pittier	Clay
Fabaceae (Faboid)	*Swartzia cardiosperma* Spruce ex Benth.	White sand
Fabaceae (Mimosoid)	*Parkia multijuga* Benth.	Clay
Fabaceae (Mimosoid)	*Parkia igneiflora* Ducke	White sand

Using a 21-month reciprocal transplant experiment with 20 common white-sand and clay forest species from six genera, and manipulating the presence of herbivores, we directly assessed the effect of herbivory on the maintenance of habitat specialization in trees. We addressed three main questions: (i) Can soil specialists grow in a different edaphic environment, or is soil type alone a sufficient barrier for these plants? (ii) Do herbivores exclude clay specialists from white-sand forests? (iii) Do white-sand specialists grow more slowly than clay forest specialists, putting them at a competitive disadvantage in clay forests?

We transplanted 880 seedlings from six genera arranged in pairs of clay/white-sand specialist congeners (Fig. 1). In May 2001, we built 22 control and 22 herbivore exclosures (3 m by 3 m by 2 m) in clay and white-sand habitats, controlling for light lev-

Fig. 2. The effects of habitat and herbivore protection on (**A**) leaf area growth rate, (**B**) meristem height growth rate, and (**C**) percent mortality for white-sand and clay specialist species. Bars represent mean and ±1 SE. Values with different letters (a, b, and c) are significantly different from one another [Tukey tests for (A) and (B); Mann-Whitney U for (C)].

els (25). One seedling of each species was transplanted into each of the 22 controls and 22 exclosures, for a total of 20 seedlings per exclosure. The exclosures were completely covered by 1-mm fine nylon netting, whereas the controls had netting only on the roof. The experiment lasted until February 2003. To assess differences between seedlings in white-sand versus clay forests, and in control versus protected treatments (exclosures), we measured growth rate (leaf area per day and meristem height per day) and mortality (25).

Both clay and white-sand specialist species exhibited positive growth on both soil types. In white-sand habitat, clay specialists and white-sand specialists suffered statistically similar mortality when protected from herbivores (Fig. 2). These findings contradict the hypothesis that white-sand soil is probably toxic to nonadapted edaphic specialists because of differences in H^+/Al ratios (26). Our results also contradict a related hypothesis that posits that the low nutrient availability and/or water retention capacity of white sand exclude most nonspecialist plants from white-sand forests (6, 27, 28). Our results showed that when protected from herbivores and planted in white-sand habitat, clay species produced more than twice the amount of leaf area [0.21 cm^2 day^{-1} versus 0.09 cm^2 day^{-1}, $P < 0.05$ (Fig. 2A)] and grew significantly taller than white-sand species [0.09 mm day^{-1} versus 0.05 mm day^{-1}, $P < 0.05$ (Fig. 2B)]. Summing the 18 months of measurements, protected clay specialists grew an average of 4.9 cm in height and 115.3 cm^2 in leaf area as compared to protected white-sand specialists' 2.7 cm in height and 49.4 cm^2 in leaf area. Although soil type does have a significant effect on seedling meristem growth rate ($P < 0.05$, table S3), clearly the restriction of clay specialist species to clay habitat cannot be explained solely as a function of soil type, at least not at the critical seedling stage.

Our results showed a strong effect of herbivores on the exclusion of clay specialist species from white-sand forests. When unprotected from herbivores in white-sand habitat, white-sand specialists exhibited a significantly higher leaf growth rate than clay specialists [0.15 cm^2 day^{-1} versus −0.4 cm^2 day^{-1}, $P < 0.05$ (Fig. 2A)] and suffered only half the mortality rate [9% versus 20%, $P < 0.05$ (Fig. 2C and fig. S2)]. Taken together, the results from white-sand forests indicate that without the presence of herbivores, clay specialists survived just as well as white-sand specialists but grew taller and produced more leaf area and thus potentially could invade and displace white-sand specialists from white-sand forests.

Furthermore, white-sand specialist and clay specialist species responded differently to herbivore protection in both leaf and mer-

istem growth rate [origin × protection interaction, $P < 0.0001$ and $P < 0.01$, respectively (tables S2 and S3)] and in mortality rate ($P < 0.05$) (Fig. 2). White-sand species did not respond to protection in either habitat ($P > 0.05$), presumably because their leaves are already highly defended against herbivores (23, 29). Indeed, chemical analyses of these seedlings showed that leaves of white-sand specialists contain significantly higher tannin: protein ratios than leaves of clay specialists (30). In contrast, clay specialist species gained a significant advantage from herbivore protection, producing 267% more leaf area per day when protected (Fig. 2A).

All species grew better on clay soils, but clay specialists outperformed white-sand specialists on clay soils, providing indirect evidence that interspecific competition limits the invasion of clay forests by white-sand species. In the clay soil, white-sand species exhibited significantly slower height ($P < 0.01$) and leaf ($P < 0.01$) growth than clay specialists [habitat × origin interaction (Fig. 2, A and B, and tables S2 and S3)]. These results are consistent with the hypothesis that there is a growth/defense trade-off in these tree seedlings and that the high defense levels of white-sand specialists preclude a rapid growth rate and place them at a competitive disadvantage in clay soils.

We conclude that herbivores play an important role in maintaining habitat specialization in clay and white-sand specialist trees. Although other factors certainly influence habitat specialization at other stages of a tree's life (such as seed predation, the ability to tolerate falling debris, root morphology, etc.), herbivores appear to have primary importance at the seedling stage, and in addition, likely continue to attack trees throughout a tree's life. Herbivores killed a significant percentage of tree seedlings in this system, and in white-sand forests they selectively attacked clay specialist transplants, maintaining the dominance of endemic white-sand specialists. Without herbivores, our results indicate that clay specialists outperform white-sand specialists in both soil types. However, with herbivores present, clay specialist species dominate only in clay forests and white-sand specialists dominate only in white-sand forests. The strong negative impact of herbivores on the growth and mortality rates of clay specialists underscores the intense selective pressure in white-sand forests favoring a large investment in antiherbivore defenses. Thus, the trade-off between growth rate and antiherbivore defense appears to restrict species to one habitat or the other for the six genera we studied, and habitat specialization results from an interaction of herbivore pressure with soil type.

Edaphic heterogeneity, especially in the tropics, has been implicated in the generation
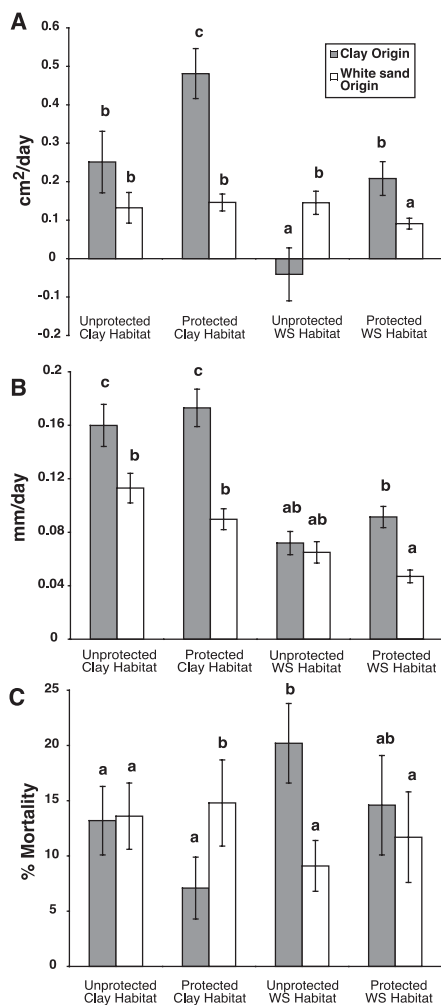
and maintenance of high diversity (*1–4, 20, 31, 32*), even though the specific mechanisms by which edaphic factors might influence diversity have, until now, never been tested (*33*). Our results suggest a mechanism by which herbivores may influence plant evolution: by accentuating habitat differences and thereby increasing the potential for edaphic heterogeneity to produce habitat specialization. There is mounting evidence that parapatric speciation across environmental gradients can occur in the face of gene flow (*34–36*). Thus, our data suggest that herbivores can be viewed as a diversifying force because they make existing abiotic gradients more divergent and cause finer-scale habitat specialization by magnifying the differences between habitats. Our study highlights the importance of edaphic differences, but the same mechanism could work across other abiotic gradients such as altitude, rainfall, and flooding regime—wherever the impact of herbivores is dissimilar across a habitat boundary.

References and Notes

1. A. H. Gentry, *Ann. MO Bot. Gard.* **75**, 1 (1988).
2. D. B. Clark, M. W. Palmer, D. A. Clark, *Ecology* **80**, 2662 (1999).
3. H. Tuomisto *et al.*, *Science* **269**, 63 (1995).
4. H. Tuomisto, K. Ruokolainen, M. Yli-Halli, *Science* **299**, 241 (2003).
5. J. L. Harper, *The Population Biology of Plants* (Academic Press, New York, 1977).
6. D. A. Coomes, P. J. Grubb, *Vegetatio* **122**, 167 (1996).
7. A. Mark, J. Dickinson, J. Allen, R. Smith, C. West, *Austral Ecol.* **26**, 423 (2001).
8. S. Louda, K. H. Keeler, R. D. Holt, in *Perspectives in Plant Competition*, J. Grace, D. Tilman, Eds. (Academic Press, New York, 1990), pp. 413–444.
9. N. S. Cobb *et al.*, *Oecologia* **109**, 389 (1997).
10. H. Olff, M. Ritchie, *Trends Ecol. Evol.* **13**, 261 (1998).
11. M. Parker, R. Root, *Ecology* **62**, 1390 (1981).
12. S. M. Louda, *Ecol. Monogr.* **52**, 25 (1982).
13. S. M. Louda, *Ecology* **64**, 511 (1983).
14. S. M. Louda, J. Rodman, *J. Ecol.* **84**, 229 (1996).
15. C. D. G. Harley, *Ecology* **84**, 1477 (2003).
16. S. J. DeWalt, J. S. Denslow, K. Ickes, *Ecology* **85**, 471 (2004).
17. R. T. King, *Biotropica* **35**, 462 (2003).
18. P. D. Coley, J. A. Barone, *Annu. Rev. Ecol. Syst.* **27**, 305 (1996).
19. K. Ruokolainen, H. Tuomisto, in *Geoecologia y Desarrollo Amazonico: Estudio Integrado en la Zona de Iquitos, Peru*, R. Kalliola, S. Flores, Eds. (Univ. of Turku Press, Turku, Finland, 1998), pp. 253–368.
20. A. H. Gentry, in *Conservation Biology: The Science of Scarcity and Diversity*, M. Soulé, Ed. (Sinauer, Sunderland, MA, 1986), pp. 153–181.
21. K. Ruokolainen, A. Linna, H. Tuomisto, *J. Trop. Ecol.* **13**, 243 (1997).
22. R. Vásquez M., *Florula de las Reservas Biologicas de Iquitos, Peru* (Missouri Botanical Garden, St. Louis, MO, 1997).
23. D. H. Janzen, *Biotropica* **6**, 69 (1974).
24. P. D. Coley, J. P. Bryant, F. S. Chapin, *Science* **230**, 895 (1985).
25. Detailed information on materials and methods is available on *Science* Online.
26. J. Proctor, *Bot. J. Scotl.* **51**, 1 (1999).
27. E. Medina, E. Cuevas, in *Mineral Nutrients in Tropical Forest and Savanna Ecosystems*, J. Proctor, Ed. (Blackwell, Oxford, 1989), pp. 217–240.
28. J. A. Moran, M. G. Barker, A. J. Moran, P. Becker, S. M. Ross, *Biotropica* **32**, 2 (2000).
29. D. McKey, P. G. Waterman, N. Mbi, J. S. Gartlan, T. T. Struhsaker, *Science* **202**, 61 (1978).
30. P. V. A. Fine *et al.*, unpublished data.
31. K. Young, B. León, *Brittonia* **41**, 388 (1989).
32. R. Condit *et al.*, *Science* **299**, 666 (2002).
33. P. Sollins, *Ecology* **79**, 23 (1998).
34. J. A. Endler, *Geographic Variation, Speciation, and Clines* (Princeton Univ. Press, Princeton, NJ, 1977).
35. C. Moritz, J. Patton, C. Schneider, T. B. Smith, *Annu. Rev. Ecol. Syst.* **31**, 533 (2000).
36. R. Ogden, R. S. Thorpe, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 13612 (2002).
37. M. W. Chase *et al.*, *Ann. MO Bot. Gard.* **80**, 528 (1993).
38. P. V. A. Fine, D. C. Daly, G. Villa, I. Mesones, K. M. Cameron, unpublished data (available on request from the first author).
39. R. García-Villacorta, M. Ahuite R., M. Olórtegui Z., *Folia Amazónica* **14**, 11 (2002).
40. P. Jørgensen, S. Léon-Yánez, *Catalogue of the Vascular Plants of Ecuador* (Missouri Botanical Garden Press, St. Louis, MO, 1999).
41. We thank the Peruvian Ministry of Natural Resources (INRENA) for permission to conduct this study; D. Del Castillo, L. Campos, E. Rengifo, and S. Tello of the Instituto de Investigaciones de la Amazonía Peruana

Silencing the Jasmonate Cascade: Induced Plant Defenses and Insect Populations

André Kessler, Rayko Halitschke, Ian T. Baldwin*

We transformed the native tobacco, *Nicotiana attenuata*, to silence its lipoxygenase, hydroperoxide lyase, and allene oxide synthase genes in order to inhibit oxylipin signaling, known to mediate the plant's direct and indirect defenses. When planted into native habitats, lipoxygenase-deficient plants were more vulnerable to *N. attenuata*'s adapted herbivores but also attracted novel herbivore species, which fed and reproduced successfully. In addition to highlighting the value of genetically silencing plants to study ecological interactions in nature, these results show that lipoxygenase-dependent signaling determines host selection for opportunistic herbivores and that induced defenses influence herbivore community composition.

The plant traits that are important for resistance to herbivore attack in nature are complex and operate on many spatial scales. They involve direct defenses (toxins, digestibility reducers, etc.) (*1*), which themselves protect plants, as well as indirect defenses, which work with components of a plant's community (natural enemies, diseases, etc.) (*2–6*). Moreover, plant defenses can be constitutively expressed or produced in response to an attacking pathogen or herbivore, when they are needed (*7, 8*).

Phenotypic plastic responses such as herbivore-induced plant defenses are thought to be an adaptation to unpredictable environments (*8*). In native populations of *Nicotiana attenuata* in the southwestern United States, herbivory is an unpredictable selective factor. *N. attenuata*'s ephemeral occurrence after fires (*9*) forces its herbivore community to

reestablish itself with every new plant population, and the plant produces a wide array of direct and indirect chemical defenses in response to this unpredictable herbivore attack (*10*). Many of the responses are specifically elicited by signals introduced into wounds during feeding (*11*), and most herbivore-induced responses studied to date require oxylipin signaling (*12, 13*).

Genetic transformation provides a valuable tool with which to manipulate traits that mediate complex plant-herbivore interactions and allows an integrative analysis of single traits (*14–16*). However, transformants are usually evaluated with known challenges, not the vast number of unknowns that occur in nature. We used transformed lines of the wild tobacco species *N. attenuata*, which express *N. attenuata* lipoxygenase 3 (*NaLOX3*), hydroperoxide lyase (*NaHPL*), and allene oxide synthase (*NaAOS*) in an antisense orientation (as-*lox*, as-*hpl*, as-*aos*, respectively) (*17*) to study herbivore-induced plant responses in nature. All three enzymes are key regulators in two distinct oxylipin pathways (fig.

Department of Molecular Ecology, Max-Planck-Institute for Chemical Ecology, Hans-Knöll-Strasse 8, Jena 07745, Germany.

*To whom correspondence should be addressed. E-mail: baldwin@ice.mpg.de

S1A) and are known to play a major role in plants' wound recognition and signaling (*13*), but their defensive function in the rough-and-tumble of the natural environment is unknown.

In laboratory studies, plants deficient in the expression or recognition of jasmonates (JAs) derived from LOX3 are unable to elicit defense compounds and are more susceptible to herbivore attack (*15*, *18*, *19*). Similarly, silencing AOS in *N. attenuata* plants partially reduces JA and defense compound accumulation but does not attenuate the resistance to herbivores (*20*). HPL-derived C_6-aldehydes and -alcohols (green leaf volatiles, or GLVs) are antimicrobial and can function as direct defenses against some herbivores (*21*). Moreover, some C_6-compounds may function as indirect defenses (*5*) or play a role in eliciting defense gene expression (*22*) and signaling within or between plants (*23*). However, the function of oxylipins is not restricted to wound signaling and defense induction but includes the regulation of tuber growth (*15*), trichome (*19*) and flower development (*24*, *25*), and ultraviolet B protection. Thus, how LOX3-, AOS-, and HPL-deficient *N. attenuata* plants would behave in their natural environment with its complexity of stressors was unknown. To extend the laboratory characterization of these genetically transformed plant lines (*18*, *20*), we examined their growth and induced production of volatile organic compounds (VOCs), and evaluated their resistance to a specialized herbivore (*Manduca sexta*) under field conditions. We transplanted young rosette plants into the field plots and allowed them to establish for at least 1 week before experimenting (*17*). The growth rates [analysis of variance (ANOVA), $F_{3,118} = 1.837$, $P = 0.144$] and morphology of the three transformed plant lines did not differ from those of wild-type (WT) plants under the noncompetitive conditions of the experiment (*26*). The plants, however, did differ in their production of VOCs (*17*).

The release of herbivore-induced terpenoid compounds, in particular the sesquiterpene, *cis*-α-bergamotene, is a demonstrated sentinel for other herbivore-induced changes in oxylipin-dependent secondary metabolite production of *N. attenuata* plants (*10*, *17*, *18*). Undamaged plants of all four genotypes released equally small amounts of *cis*-α-bergamotene (ANOVA, $F_{3,11} = 0.876$, $P = 0.483$). WT plants and HPL-deficient plants released significantly increased amounts of *cis*-α-bergamotene and a suite of other terpenoid compounds after *M. sexta* damage (table S1). In contrast, the *cis*-α-bergamotene emission from herbivore-damaged as-*lox* and as-*aos* (ANOVA, $F_{3,10} = 44.56$, $P < 0.0001$; Bonferroni post hoc $P < 0.05$) plants remained low (Fig. 1A).

In addition to the open-flow VOC trapping design, we used a portable gas chromatograph (z-Nose) (*27*) to characterize the wound-induced emissions of the GLV, *cis*-3-hexenal, from WT and transformed plants (*17*). As-*hpl* plants released significantly less *cis*-3-hexenal than did WT, as-*aos*, and as-*lox* lines immediately after damaging the leaf tissues (ANOVA, $F_{3,19} = 33.07$, $P < 0.0001$; Bonferroni post hoc $P < 0.05$) (Fig. 1B). This demonstrates that the production of GLVs, such as *cis*-3-hexenal, requires the activity of HPL but not LOX3, which specifically supplies fatty acid hydroperoxides to the octadecanoid pathway but not to the HPL reaction (*17*, *18*) (fig. S1). Interestingly, AOS-deficient plants had significantly higher *cis*-3-hexenal emission levels after the mechanical damage than did all other plant lines (Bonferroni post hoc $P < 0.05$) (Fig. 1B), which suggests either a rechanneling of AOS substrates into the HPL cascade (*20*) and/or an octadecanoid-derived negative regulation of the HPL cascade.

The altered production of secondary metabolites, such as terpenoids, in as-*lox* plants was correlated with reduced resistance to attack by the specialist herbivore *M. sexta*. Freshly hatched *M. sexta* caterpillars, which fed on rosette-stage, field-grown WT plants and similar-sized plants of the three different transformed lines (*17*), gained weight fastest on LOX3-deficient plants (ANOVA, $F_{3,23} = 5.83$, $P = 0.0041$). After 9 days of development on as-*lox* plants, they were 4.4-fold heavier than caterpillars on WT plants and 7-fold and 2.5-fold heavier than those on HPL- and AOS-deficient plants (Bonferroni post hoc tests, $P < 0.05$), respectively (Fig. 1C). In laboratory experiments, herbivore resistance can be restored by treating as-*lox* plants with methyl jasmonate (*20*). The relative resistance of AOS-deficient plants to herbivores and the susceptibility of LOX3-deficient plants confirm laboratory results and correlated with the wound-induced production of nicotine. Nicotine production was reduced only in as-*lox* plants, which was attributed to a leaky phenotype in the as-*aos* plants (fig. S1B) (*17*, *20*). Similarly, as-*hpl* plants retain their resistance to hornworm damage despite the evidence that HPL-derived oxylipins may signal plant-defense activation (*22*, *23*). In laboratory studies, hornworm consumption and growth were slower on HPL-deficient plants than on WT plants but could be restored to WT levels by the addition of GLVs to as-*hpl* plants, which suggested that GLVs stimulate feeding by *M. sexta* (*20*).

The initial field experiments showed that the three transformed plant lines (as-*lox*, as-*aos*, as-*hpl*) have similar characteristics in the field and in the laboratory (*18*, *20*) and differ in their responsiveness to *M. sexta* attack. However, the herbivore community of *N. attenuata* is diverse and includes piercing-sucking, leaf-mining, and leaf-chewing herbivores, such as *M. sexta*. Moreover, the composition of the herbivore community can be extremely variable and is influenced by random effects (*5*, *9*, *10*). Therefore, the exposure of genetically manipulated plants to their natural herbivore community realistically evaluates the role played by oxylipin-mediated herbivore-induced responses in structuring the plants' herbivore community.
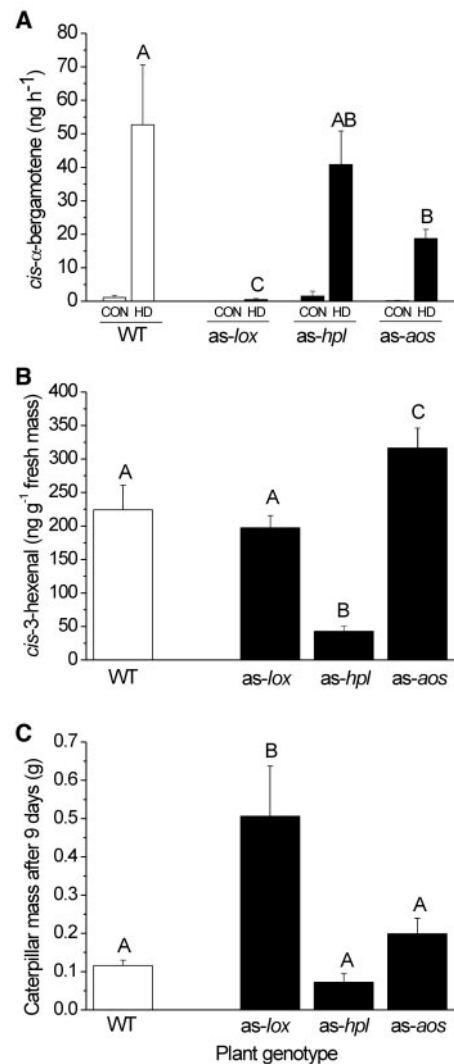

Fig. 1. Volatile organic compound emission and herbivore susceptibility of *N. attenuata* wild-type (WT) plants and plants transformed to silence LOX3 (as-*lox*), HPL (as-*hpl*), and AOS (as-*aos*) activity. (**A**) Mean (+SEM) *cis*-α-bergamotene emission (ng hour^{-1}) from undamaged (CON) and *Manduca sexta* hornworm-damaged (HD) plants of the four genotypes. (**B**) Mean (+SEM) emission of *cis*-3-hexenal (ng g^{-1} fresh mass) measured with a portable gas chromatograph (z-Nose) in the headspace of excised leaves that had been mechanically wounded. (**C**) Mean (+SEM) caterpillar mass (g) after 9 days of development on WT plants compared to the genetically transformed lines. Different letters designate significantly different means as informed by a Bonferroni post hoc ($P < 0.05$) test of an ANOVA.

We transplanted 80 young rosette-stage plants in groups of four into a field plot along a linear transect (*17*). Each group comprised one WT plant and one plant of each of the transformed lines: as-*lox*, as-*hpl*, and as-*aos*. Extensive prior laboratory analysis of plants transformed with empty-vector constructs revealed no differences in any measured herbivore resistance trait or growth and reproduction when compared to WT plants (*16*). During the 2002–2003 field seasons, we observed damage from herbivores of all three feeding guilds but found very few plants that were attacked by the most damaging herbivores of previous field seasons: the leaf-chewing sphingid hornworms *Manduca quinquemaculata* and *M. sexta* and the piercing-sucking mirid bug *Tupiocoris notatus* (*5*, *10*). Thus the observed damage, though relatively low, allowed the distribution of herbivory on the silenced lines to be clearly evaluated.

If plants within a population show induced defense responses to herbivory and if herbivores can freely choose among plants, herbivory should be equally distributed. Plants lacking the ability to respond to herbivore attack are more susceptible and should be preferred to responsive plants (*7*, *8*). Consistent with this prediction, we found dramatically more herbivore damage on as-*lox* plants (ANOVA, $F_{3,120} = 24.19$, $P < 0.0001$; Bonferroni post hoc $P < 0.05$) than on WT, as-*hpl*, or as-*aos* plants (Fig. 2A, fig. S2A), which corresponds to the increased performance of *M. sexta* caterpillars on as-*lox* plants. Similarly, herbivores attacked a significantly greater proportion of as-*lox* plants compared to all other lines in the experimental population (Fig. 2A). Although herbivory was equally distributed among WT, as-*aos*, and as-*hpl* plants (Pearson chi-square test $\chi^2 < 2.518$, $P > 0.05$), it was significantly higher on plants with strongly attenuated induced responses (as-*lox*) (Pearson chi-square test $\chi^2 > 49.207$, $P < 0.0001$), which suggests that a plant's ability to elicit defenses influences the distribution of herbivory within a plant population.

Moreover, a more detailed analysis of the herbivore community revealed that the herbivore-induced plant responses can alter the host breadth of generalist herbivores. Compared to previous study years, we found two new herbivores on the experimental plants: the leafhopper *Empoasca* sp. (*28*) (Fig. 3, fig. S2B) and the western cucumber beetle *Diabrotica undecimpunctata tenella* Le Conte (fig. S2C). In fact, most of the observed damage on LOX3-deficient plants resulted from just one of these, *Empoasca*
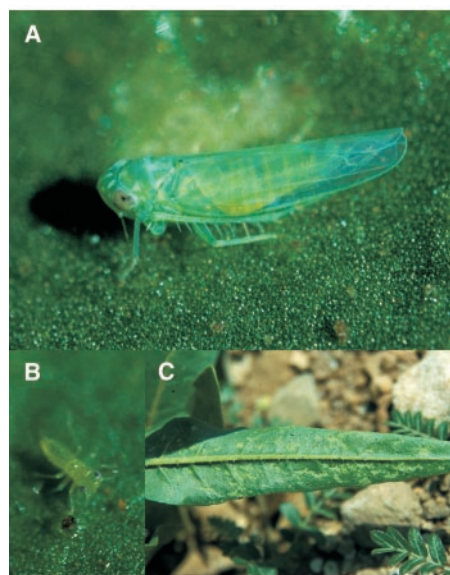
sp., which is an unusual herbivore on *N. attenuata* plants. In 12 quantitative formal surveys of 600 plants conducted over four consecutive years, we found a total of only four *Empoasca* sp. specimens on *N. attenuata* plants growing in native populations (table S2). Neither in these surveys, nor in informal observations of more than 40,000 plants over the past 15 years of field work has there been evidence of *Empoasca* feeding damage. *Empoasca* sp. is a highly mobile, opportunistic herbivore that attacks *Datura wrightii*, *Solanum americanum*, and *Mirabilis multiflora*, all of which were abundant in the study area. In our field trials with different *N. attenuata* lines, the as-*lox* plants received significantly more *Empoasca* damage than did the other transformed lines (as-*aos*, as-*hpl*) or the WT (ANOVA, $F_{3,121} = 30.25$, $P < 0.0001$, Bonferroni post hoc test $P < 0.05$) (Fig. 2B). More than 68% of the as-*lox* plants in the experimental population were damaged by *Empoasca*, whereas the proportion of damaged plants in the other lines did not exceed 29% (Fig. 2B). More importantly, *Empoasca* females oviposited on as-*lox* plants, and these plants were susceptible to attack from the emerging leafhopper offspring (Fig. 3B). *Empoasca* nymphs were found on 81% of the attacked as-*lox* plants but not at all on WT, HPL-, and AOS-deficient lines. Moreover, when given the choice (*17*), adult leafhoppers clearly prefer as-*lox* plants over WT plants (paired Student's *t* test, $t = -4.919$, $P = 0.0012$) (Fig. 4). Interestingly, we found no *Empoasca* damage on as-*hpl* plants, suggesting that leafhopper feeding behavior may also be influenced by GLVs, as has been demonstrated with *M. sexta* larvae (*20*).

The reduced resistance of LOX3-deficient plants was correlated with an altered octadecanoid-dependent gene expression in response to *Empoasca* attack. With a cDNA microarray analysis, we compared transcript accumulation in response to *Empoasca* feeding in WT and as-*lox* plants to identify
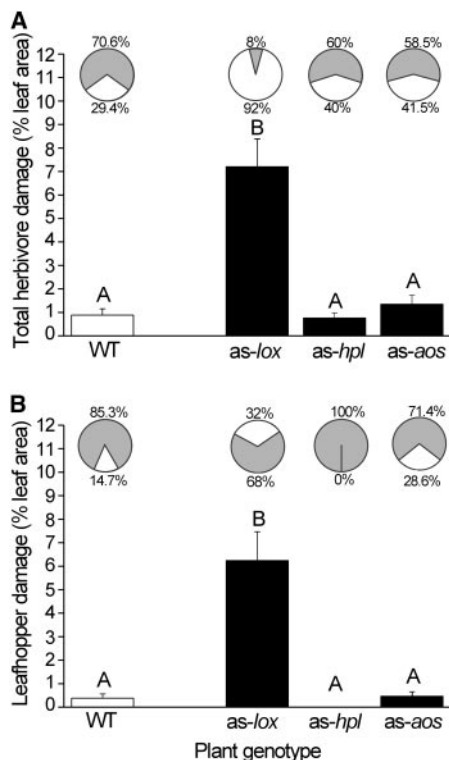
Fig. 2. Herbivore damage on field-grown *N. attenuata* wild-type (WT) plants and plants transformed to silence LOX3 (as-*lox*), HPL (as-*hpl*), and AOS (as-*aos*) activity. (**A**) Mean (+SEM) percentage of leaf area damaged by herbivores (bar graph) and proportion of plants in the population that were attacked by herbivores (white in the pie chart) or remained undamaged (gray in the pie chart). (**B**) Mean (+SEM) percentage of leaf area damaged by *Empoasca* sp. leafhoppers (bar graph) and proportion of plants in the population that were attacked by leafhoppers (white in the pie chart) or remained undamaged (gray in the pie chart). Different letters designate significantly different means as informed by a Bonferroni post hoc ($P < 0.05$) test of an ANOVA.

Fig. 3. *Empoasca* sp. leafhoppers on as-*lox* *N. attenuata* plants. (**A**) Adult *Empoasca* sp. (**B**) *Empoasca* sp. nymphs were exclusively found on as-*lox* plants. (**C**) Visual damage by *Empoasca* sp. on as-*lox* *N. attenuata* plants.

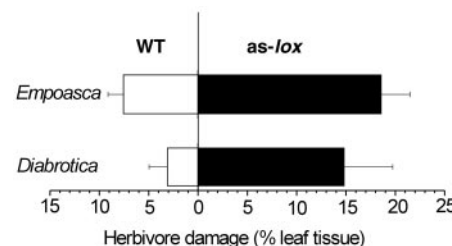**Fig. 4.** Mean (+SEM) leaf tissue damage on wild-type (WT) and as-*lox* plants after 3 days of attack by *Empoasca* sp. leafhoppers or *D. undecimpunctata* leaf beetles, respectively. Ten *Empoasca* and two *D. undecimpunctata*, respectively, were allowed to choose between a WT and an as-*lox* plant that had been potted together and covered with insect mesh. Both experiments were replicated 10 times.

octadecanoid-dependent transcriptional responses (fig. S3) (*17*). Of the 240 *N. attenuata* genes spotted on the microarray, 54 genes showed significantly increased LOX3-dependent transcript accumulation after *Empoasca* damage and 50 showed decreased accumulation (table S3). A series of defense-related genes, such as trypsin proteinase inhibitor (TPI) and threonine deaminase, had increased expression levels, whereas others, such as xyloglucan endotransglucosylase/hydrolase or GAL83 and many of the photosynthesis-related genes (e.g., small subunit of ribulose bisphosphate carboxylase), were down-regulated in response to *Empoasca* attack and showed little if any regulation in as-*lox* plants (table S3). These results suggest that a complex LOX3-dependent regulation of primary and secondary metabolism in *N. attenuata* mediates resistance to the piercing-sucking *Empoasca* leafhoppers. The fact that octadecanoid signaling plays a role in plant defense against both piercing-sucking and leaf-chewing insects suggests a common response to attack from members of these two feeding guilds. Moreover, LOX3-dependent octadecanoids may play a direct role in host-plant selection by enabling herbivores to differentiate between plants with and without intact JA signaling, as shown in experiments with the corn earworm, *Helicoverpa zea*. This generalist herbivore uses induced JA and salicylate to activate four of its cytochrome P450 genes that are associated with detoxification either before or concomitantly with the biosynthesis of allelochemicals (*29*).

In addition to *M. sexta*, which was affected by LOX3-mediated plant resistance traits, we found a novel leaf-chewing herbivore on *N. attenuata*, the leaf beetle *Diabrotica undecimpunctata tenella* Le Conte (fig. S2C). It often feeds on *D. wrightii* and *Cucurbita foetidissima* flowers in the study area and was observed on *N. attenuata* plants exclusively in this study and only on as-*lox* plants. To test whether or not this clear preference in the field is caused by the decreased expression of LOX3, we allowed the beetles to choose between WT and as-*lox* plants. The choice experiment revealed a clear preference for as-*lox* plants compared to WT plants (paired Student's *t* test, $t = -4.050$, $P = 0.003$) (Fig. 4).

Our results show that the LOX3-mediated inducibility of plants is crucial for the oviposition decision and for the opportunistic host selection behavior of *Empoasca* sp. and *D. undecimpunctata*, and thereby defines host breadth. Therefore, host selection seems determined not only by the plant's constitutively expressed chemical phenotype and external mortality factors (predation pressure, abiotic stress) (*30*) but also by the plant's ability to induce responses to herbivory. The

additional finding that induced responses to herbivory influence the distribution of herbivory within a plant community points to the value of genetically silenced plants in ecological research. An understanding of the ecological interactions that occur in nature is essential for sustainable agriculture.

References and Notes

1. S. S. Duffey, M. J. Stout, *Arch. Insect Biochem. Physiol.* **32**, 3 (1996).
2. C. M. De Moraes, W. J. Lewis, P. W. Pare, H. T. Alborn, J. H. Tumlinson, *Nature* **393**, 570 (1998).
3. M. Dicke, J. J. A. van Loon, *Entomol. Exp. Appl.* **97**, 237 (2000).
4. J. S. Thaler, M. J. Stout, R. Karban, S. S. Duffey, *Ecol. Entomol.* **26**, 312 (2001).
5. A. Kessler, I. T. Baldwin, *Science* **291**, 2141 (2001).
6. T. C. J. Turlings, J. H. Tumlinson, W. J. Lewis, *Science* **250**, 1251 (1990).
7. R. Karban, I. T. Baldwin, *Induced Responses to Herbivory* (Univ. of Chicago Press, Chicago, IL, 1997).
8. A. A. Agrawal, *Science* **294**, 321 (2001).
9. C. A. Preston, I. T. Baldwin, *Ecology* **80**, 481 (1999).
10. A. Kessler, I. T. Baldwin, *Plant J.* **38**, 639 (2004).
11. R. Halitschke, U. Schittko, G. Pohnert, W. Boland, I. T. Baldwin, *Plant Physiol.* **125**, 711 (2001).
12. E. Blee, *Trends Plant Sci.* **7**, 315 (2002).
13. R. A. Creelman, J. E. Mullet, *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **48**, 355 (1997).
14. J. Bergelson, C. B. Purrington, C. J. Palm, J. C. Lopez-Gutierrez, *Proc. R. Soc. London B Biol. Sci.* **263**, 1659 (1996).
15. J. Royo et al., *Proc. Natl. Acad. Sci. U.S.A.* **96**, 1146 (1999).
16. J. A. Zavala, A. G. Patankar, K. Gase, I. T. Baldwin, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 1607 (2004).
17. Materials and methods are available as supporting material on *Science* Online.
18. R. Halitschke, I. T. Baldwin, *Plant J.* **36**, 794 (2003).
19. L. Li et al., *Plant Cell* **16**, 126 (2004).
20. R. Halitschke, M. Keinanen, J. Ziegler, I. T. Baldwin, *Plant J.*, in press.
21. D. F. Hildebrand, G. C. Brown, D. M. Jackson, T. R. Hamilton-Kemp, *J. Chem. Ecol.* **19**, 1875 (1993).
22. N. J. Bate, S. J. Rothstein, *Plant J.* **16**, 561 (1998).
23. G. Arimura, R. Ozawa, J. Horiuchi, T. Nishioka, J. Takabayashi, *Biochem. Syst. Ecol.* **29**, 1049 (2001).
24. J. H. Park et al., *Plant J.* **31**, 1 (2002).
25. A. Stinzi, J. Browse, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 10625 (2000).
26. As plants began to elongate and produce flowers, they were examined daily; all flowers were removed before opening and anthesis to meet the performance standards determined in the Code of Federal Regulations [7CFR340.3(c)] for the Introduction of Organisms Altered or Produced through Genetic Engineering. Consequently, direct fitness measures were unattainable in these experiments.
27. M. Kunert, A. Biedermann, T. Koch, W. Boland, *J. Sep. Sci.* **25**, 677 (2002).
28. Probably belonging to the species *E. fabae* (Harris, 1841), which is widely known as a pest on alfalfa in the United States.
29. X. C. Li, M. A. Schuler, M. R. Berenbaum, *Nature* **419**, 712 (2002).
30. E. Bernays, M. Graham, *Ecology* **69**, 886 (1988).
31. Supported by the Max-Planck-Gesellschaft and Deutsche Forschungsgemeinschaft (BA-2132/1-1). We thank E. Wheeler, E. Pichersky, D. Heckel, J. Gershenzon, M. Heil, and H. Vogel for helpful comments; R. Baumann, P. Freytag, and H. Nickel for assistance with species determination; and Brigham Young University for use of Lytle Preserve as a field station; L. Rausing for helping us promote the discussion of the scientific value of transformed plants; and J. White and the Animal and Plant Health Inspection Service personnel for facilitating their safe use in nature.

Osedax: Bone-Eating Marine Worms with Dwarf Males

G. W. Rouse,[1,2]* S. K. Goffredi,[3]* R. C. Vrijenhoek[3]†

We describe a new genus, *Osedax*, and two new species of annelids with females that consume the bones of dead whales via ramifying roots. Molecular and morphological evidence revealed that *Osedax* belongs to the Siboglinidae, which includes pogonophoran and vestimentiferan worms from deep-sea vents, seeps, and anoxic basins. *Osedax* has skewed sex ratios with numerous dwarf (paedomorphic) males that live in the tubes of females. DNA sequences reveal that the two *Osedax* species diverged about 42 million years ago and currently maintain large populations ranging from 10^5 to 10^6 adult females.

Deep-sea exploration continues to reveal biological novelties (*1*) such as whale fall communities (*2*). Here, we describe remarkable polychaete annelids, *Osedax* gen. nov. (nov.), discovered in January 2002 on the bones of a gray whale carcass at 2891 m depth in Monterey Bay, California (*3*). Their conspicuous red plumes extended from most exposed portions of the whale bones (Figs. 1A and 2A). Colonies of these worms comprised two species, *Osedax rubiplumus* sp. nov. and *O.*

frankpressi sp. nov., that we describe along with the new genus. Nucleotide sequence analysis revealed that the two *Osedax* species differed by $17.28 \pm 0.21\%$ ($\bar{x} \pm$ SD) for mitochondrial COI, by $7.63 \pm 0.46\%$ for mitochondrial 16S rRNA, and by $4.09 \pm 0.04\%$ for nuclear 18S rRNA (*4*). On the basis of a molecular clock calibrated for COI in deep-sea annelids (*5*), *O. rubiplumus* and *O. frankpressi* diverged about 42 million years ago (Ma) (*4*), in the late Eocene, when their

ancestor may have exploited the bones of archeocete whales such as *Basilosaurus* (*6*). Phylogenetic analysis (Fig. 3) placed *Osedax* in the family Siboglinidae (*7, 8*), which includes frenulate and vestimentiferan tubeworms that also lack digestive systems (*9, 10*).

Unlike other siboglinids, female *Osedax* lack a discrete trophosome, the organ housing symbiotic bacteria in vestimentiferans and pogonophorans. Instead, *Osedax* possess a bulbous posterior ovisac covered by a sheath of green-colored tissue that branches into a vascularized "root" system and invades the bone marrow (Figs. 1, C and H, and 2, C, D, and F). This branching root system is histologically distinct and not homologous with the singular chitinous root tube found in some *Lamellibrachia* vestimentiferans (*11*). Microscopic and molecular analyses (*12*) of this sheath revealed bacteriocytes (Fig. 1J) containing large rod-shaped bacteria of the microbial order Oceanospirillales, known for heterotrophic degradation of complex organic compounds. Analyses of stable isotopes and fatty acids (*12*) revealed that the endosymbionts are responsible for the nutrition of this worm. This heterotrophic symbiosis differs markedly from the chemolithoautotrophic symbioses found in other deep-sea annelids and mollusks that rely on sulfide- or methane-oxidizing bacterial endosymbionts (*13*). This finding of an endosymbiosis involving heterotrophic degradation in *Osedax* suggests that the evolutionary history of bacterial symbioses among the Siboglinidae is more varied than previously suspected (*14*). Reliance on the bones of marine mammals, hydrocarbon degradation, and the unusual morphology of the symbiont-bearing ovisac and root system of these worms make this particular symbiotic association unique in the animal kingdom.

All *Osedax* visible to the eye were females that ranged from 0.2 to 0.5 mm in trunk width, suggesting ongoing recruitment. Females as small as 0.3 mm wide produced eggs (Figs. 1, D and F, and 2F). The tubes of individual females contain numerous microscopic males (Figs. 1, K to O, and 2G) that were filled with developing sperm (Fig. 1N) and often contained yolk droplets (Fig. 2I). These paedomorphic males retain morphological traits typical of siboglinid trochophore larvae (*15*), including a ciliary band that appears to be a putative prototroch (Figs. 1L and 2I) and opisthosomal chaetae (Figs. 1O and 2H). The tubes of larger females contained up to 111 males each, with a

male-to-female sex ratio of 17:1. Females either accumulate males over time or attract more when larger, because the number of males is correlated ($r = 0.899$, $P < 0.01$) with female size in *O. rubiplumus* (*16*). We hypothesize that sex may be environmentally determined in *Osedax*, with larvae settling on exposed bones maturing as females and those landing on females becoming males. Environmental sex determination is known in the echiuran *Bonellia* (*17*), now regarded as a polychaete (*18*).

Amounts of mitochondrial COI diversity (θ) suggest that the effective female population sizes [$N_{e(f)}$] range from 5×10^5 in *O. rubiplumus* to 9×10^5 in *O. frankpressi*.

These numbers are of the same magnitude as estimates of $N_{e(f)}$ (range of 10^5 to 10^6 females) inferred for deep-sea annelids (*4*) and consistent with estimates of $N_{e(f)}$ from other invertebrates (*19*). The large female population sizes estimated for these *Osedax* species suggest they are common on whale falls. These numbers also suggest that the frequency of whale falls has historically been great, which is consistent with estimates of large whale populations before modern whaling (*2, 20*). Their abundance suggests that these worms might play a substantial role in the cycling of large organic inputs into the surrounding deep-sea communities.
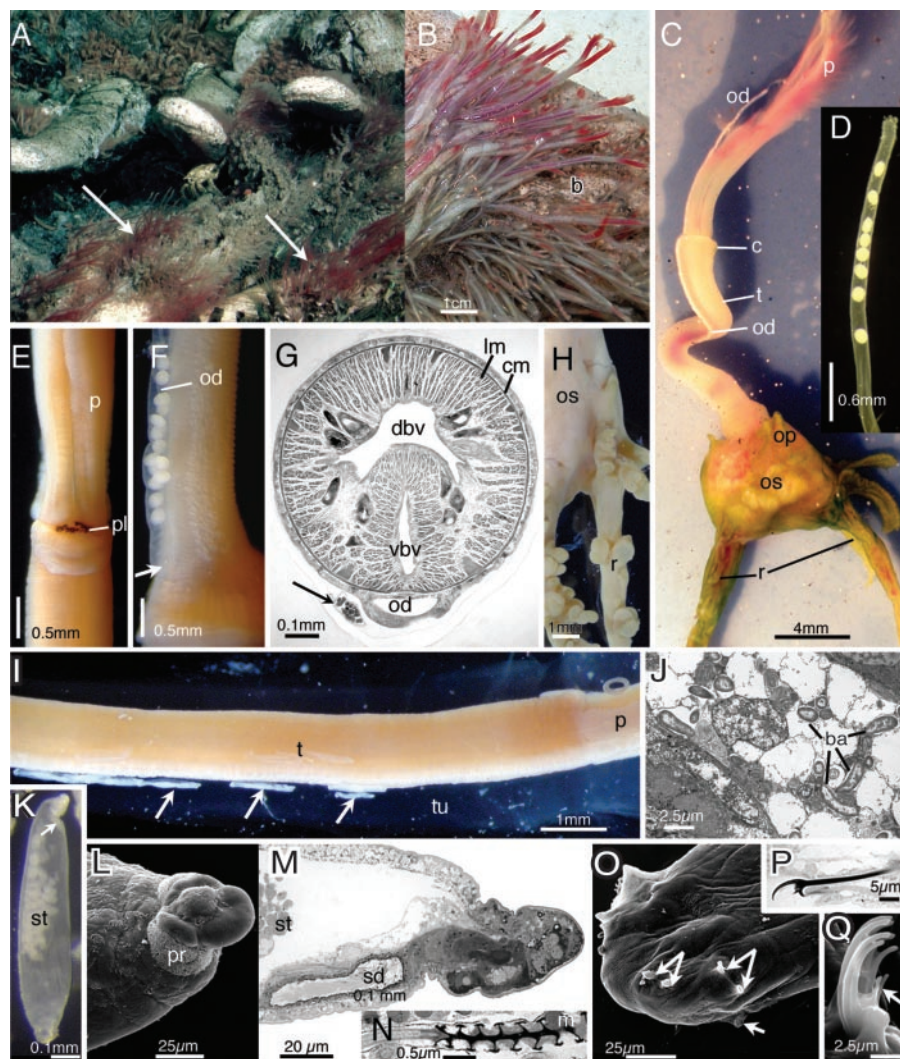

Fig. 1. *O. rubiplumus*. (**A**) Bones in situ with emergent worms (arrows). (**B**) Whale rib with female worms. (**C**) Female, tube removed, dissected from bone. (**D**) Anterior of oviduct with eggs. (**E**) Dorsal crown-trunk junction. (**F**) Lateral trunk base with oviduct emerging from ovisac (arrow). (**G**) Transverse section of anterior trunk. Male sits on oviduct (arrow). (**H**) Lateral ovisac and roots of holotype. (**I**) Dwarf males (arrows) in tube with female. (**J**) Transmission electron micrograph (TEM) of root tissue with bacteriocytes. (**K**) Dwarf male with sperm duct (arrow) and spermatids. (**L**) Scanning electron micrograph (SEM) of male anterior. (**M**) Longitudinal section (LS) through male anterior. (**N**) Longitudinal TEM section through posterior of sperm head. (**O**) SEM of posterior hooks (arrows) on male. (**P**) LS through hook. (**Q**) SEM of hook with subrostral teeth (arrow). b, bone; ba bacteria; c, collar; cm, circular muscle; dbv, dorsal blood vessel; od, oviduct; lm longitudinal muscle; m, male; op, ovisac projection; os, ovisac; p, palps; pl, plaques; pr, prototroch; r, roots; sd, sperm duct; st, spermatids; t, trunk; tu, tube; and vbv, ventral blood vessel.

[1]South Australian Museum, North Terrace, Adelaide SA 5000, Australia. [2]Earth and Environmental Sciences, University of Adelaide, Adelaide SA 5005, Australia. [3]Monterey Bay Aquarium Research Institute (MBARI), 7700 Sandholdt Road, Moss Landing, CA 95039, USA.

*These authors contributed equally to the work.
†To whom correspondence should be addressed. E-mail: vrijen@mbari.org

Systematic description. Annelida, Lamarck 1809; Canalipalpata, Rouse and Fauchald 1997; Siboglinidae Caullery, 1914. *Osedax* gen. nov. **Diagnosis.** Polychaete worms with females having a discrete red crown, contractile trunk, bulbous ovisac, and branching roots. Crown and trunk within transparent tube emergent from whale bone (Figs. 1, A and B, and 2, A to C). Crown composed of cylindrical oviduct (Figs. 1D and 2E) plus four pinnule-bearing palps (Figs. 1C and 2, D and E). No mouth or obvious gut. Cylindrical trunk composed mostly of longitudinal muscles and glands. Dorsal heart lies at anterior region of trunk; major dorsal and ventral blood vessels present (Fig. 1G). Oviduct parallels trunk surface into posterior ovisac (Figs. 1, C and F to G, and 2D) filled with numerous white oocytes. Ovisac enclosed by green sheath (Fig. 2F) composed of epidermis and bacteriocytes containing bacteria (Fig. 1J). Ovisac sheath continuous with variably branching posterior roots (Figs. 1, C and H, and 2D). Vascularized roots and ovisac (Fig. 2F). No chaetae or segmentation apparent in females. Paedomorphic males cluster around oviduct in female tubes (Figs. 1, I and K, and 2G). Males with anterior prototroch (Figs. 1L and 2I) and posterior hooked chaetae (Figs. 1O and 2H) arranged in two rows of four pairs (Fig. 1O). Hooks, lacking rostrum, comprise capitium with curved teeth over subrostral process (Figs. 1, P and Q, and 2H). Internally, males contain spermatids and sperm in anterior duct (Fig. 1, K and M, and 2I). **Etymology.** From Latin *os*, bone, and *edax*, devouring; gender masculine. **Type species.** *Osedax rubiplumus* sp. nov. by present designation.

Osedax rubiplumus sp. nov. **Type material.** (*21*) Monterey Bay, California, *Tiburon* dive T486, 36°36.8'N, 122°26.0'W, 2891 m, 8 February 2002: holotype, mature adult female (CASIZ 170238); allotypes, 38 males from tube of holotype (CASIZ 170240); paratypes, 10 females and numerous males (CASIZ 170239), 10 females and numerous males (LACM-AHF POLY 02146), and 31 females and numerous males (SAM E3376). **Diagnosis.** Holotype, emergent body in cylindrical tube, walls 1 mm thick. Contracted crown plumes 2.1 cm long. Oviduct, filled with ellipsoid eggs (mean diameters 151 μm by 121 μm, $n = 30$). Oviduct extends between palps, 1.8 cm from trunk. Palps red in living worms, pinnules on outer margins. Collar with brown plaques dorsally (Fig. 1E). Trunk 3.8 cm long, 2 mm wide at collar. Ovisac, 8 mm by 4 mm by 0.3 mm, with paired anterolateral projections (Fig. 1C) and four discrete roots posteriorly. Roots with spherical lobes (Fig. 1H). Allotypes 0.4- to 1.1-mm-long males. Putative prototroch incomplete (Fig. 1L). Posteriorly, 16 hooks with capitium teeth emergent; handles 18 to 23 μm (Fig. 1P). Capitium with six to eight teeth; two smaller subrostrumal teeth (Fig. 1Q). **Etymology.** From Latin *rubi*, red, and *pluma*, feather.

Osedax frankpressi sp. nov. **Type material.** (*21*) Monterey Bay, California, *Tiburon* dive T610, 36°36.8'N, 122°26.0'W, 2891 m, 7 August 2002: holotype, mature adult female (CASIZ 170235); allotypes, 80 males from tube of holotype (CASIZ 170237); paratypes, three females (CASIZ 170236), three females (LACM-AHF POLY 02147), three females (SAM E3377). **Diagnosis.** Holotype, emergent body in gelatinous hemispherical tube, 7-mm diameter. Contracted crown plumes 0.95 cm long. Oviduct filled with ellipsoid eggs (mean diameters 146 μm by 117 μm; $n = 15$), Oviduct convoluted upon contraction, extending between palps, 3 mm from trunk (Fig. 2, D and E). Palps red with two longitudinal white stripes in living worms (Fig. 2A); pinnules on inner margins. Trunk 4.5 mm long, 0.9 mm wide, and marked by white thickened tissue at anterior (Fig. 2E). Green, bacteriocyte-filled sheath forms trunk-ovisac junction 1.2 mm long, 1 mm wide (Fig. 2D). Lobulate ovisac, 6.5 mm by 5 mm by 3 mm (Fig. 2F). Ovisac and roots inflated with clear fluid in situ in bone (Fig. 2C); fluid lost on extraction. Allotypes, 0.15- to 0.25-mm-long males (Fig. 2, G and I). Chaetae with hooks and handles 15 to 21 μm (Fig. 2H). Capitium with five teeth; no subrostral teeth (Fig. 2H). **Etymology.** In honor of Dr. Frank Press, former U.S. presidential science advisor, president of the U.S. National Academy of Sciences, and chair of the MBARI Board of Directors, for his distinguished service to science.

Remarks. Females of the two new species are easily distinguished by the lengths of their tubes and palp coloration. *O. rubiplumus* tubes have a uniform diameter and maintain shape when removed from the water, whereas *O. frankpressi* tubes are gelatinous and collapse in air. The white-striped *O. frankpressi* palps contrast with the uniform red palps of *O. rubiplumus*. Females of the two species differ in the
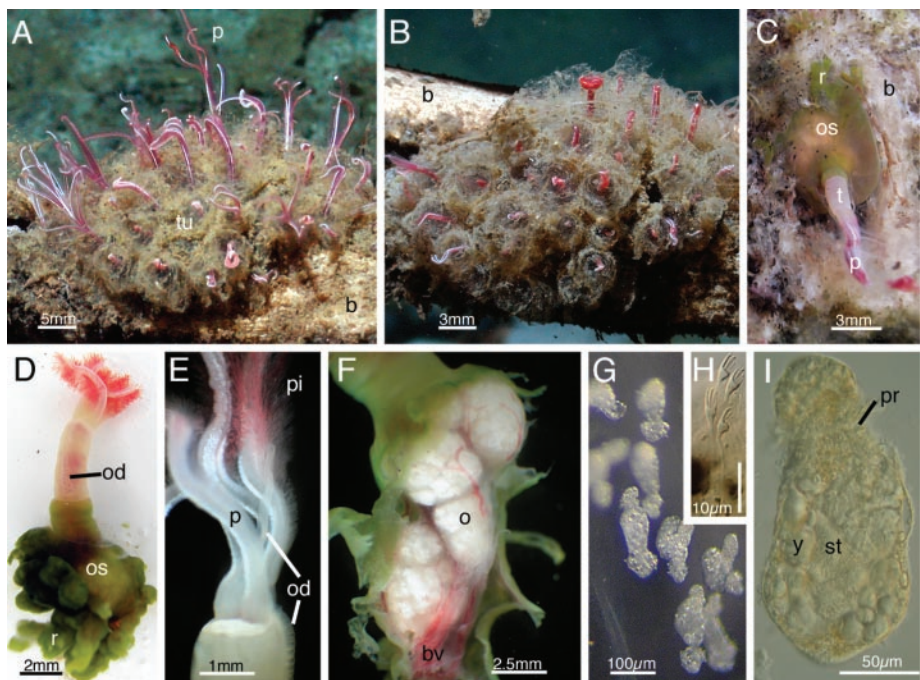
Fig. 2. *O. frankpressi*. (**A**) Whale rib in situ with emergent worms. (**B**) Retracted worms. (**C**) Female partly dissected from bone; note fluid-filled ovisac. (**D**) Female dissected from bone. (**E**) Crown-trunk junction. (**F**) Ovisac with green sheath cut to reveal ovary and blood vessels. (**G**) Dwarf males in female tube. (**H**) Male hooks. (**I**) Male showing prototroch, developing sperm, and yolk. bv, blood vessel; y, yolk; otherwise as for Fig. 1.

Fig. 3. *Osedax* as member of Siboglinidae. Bayesian analyses of 2088 molecular characters from combined 16S and 18S rDNA (*22*). Posterior probability values of 100% indicated by asterisks.

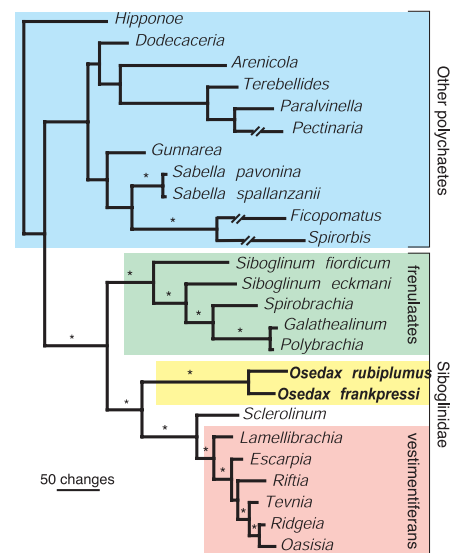trunk-ovisac junction, ovisac shape, and ramification of roots. Males of *O. frankpressi* are less than one-third the size of *O. rubiplumus* males. Male chaetae of *O. rubiplumus* have subrostral teeth and more capitium teeth.

References and Notes

1. C. L. Van Dover, C. R. German, K. G. Speer, L. M. Parson, R. C. Vrijenhoek, *Science* **295**, 1253 (2002).
2. C. R. Smith, A. R. Baco, *Oceanogr. Mar. Biol. Annu. Rev.* **41**, 311 (2003).
3. S. K. Goffredi, C. K. Paull, Fulton-Bennett, L. A. Hurtado, R. C. Vrijenhoek, *Deep Sea Res. I*, in press.
4. Details on materials and methods and GenBank accession numbers are available as supporting material on *Science* Online.
5. P. Chevaldonné, D. Jollivet, D. Desbruyères, R. A. Lutz, R. C. Vrijenhoek, *Cah. Biol. Mar.* **43**, 367 (2002).
6. P. D. Gingerich, B. H. Smith, E. L. Simons, *Science* **249**, 154 (1990).
7. K. M. Halanych, R. A. Feldman, R. C. Vrijenhoek, *Biol. Bull.* **201**, 65 (2001).
8. G. W. Rouse, *Zool. J. Linn. Soc.* **132**, 55 (2001).
9. E. C. Southward, in *Microscopic Anatomy of Invertebrates*, vol. 12 of *Onychophora, Chilopoda and Lesser Protostomata*, F. W. Harrison, M. E. Rice, Eds. (Wiley-Liss, New York, 1993), pp. 327–369.
10. M. L. Jones, *Science* **213**, 333 (1981).
11. D. Julian, F. Gaill, E. Wood, A. Arp, C. Fisher, *J. Exp. Biol.* **202**, 2245 (1999).
12. S. K. Goffredi *et al.*, in preparation.
13. C. L. Van Dover, *The Ecology of Deep-Sea Hydrothermal Vents* (Princeton Univ. Press, Princeton, NJ, 2000), p. 411.
14. A. Schulze, K. M. Halanych, *Hydrobiologia* **496**, 199 (2003).
15. E. C. Southward, *Hydrobiologia* **402**, V185 (1999).
16. Fifty-four female *Osedax rubiplumus* from a piece of rib obtained on *Tiburon* dive 486 were measured and assessed for spawning, and their tubes were checked for males. Details are available at (*4*).
17. J. Pilger, in *Settlement and Metamorphosis of Marine Invertebrate Larvae*, F.-S. Chia, M. E. Rice, Eds. (Elsevier, New York, 1978), pp. 103–112.
18. D. McHugh, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 8006 (1997).
19. M. Lynch, J. S. Conery, *Science* **302**, 1401 (2003).
20. J. Roman, S. R. Palumbi, *Science* **301**, 508 (2003).
21. Repositories of type specimens are as follows: CASIZ, California Academy of Sciences; LACM, Los Angeles County Museum; and SAM, South Australia Museum.
22. Phylogenetic analysis has its basis in 18S and 16S rDNA sequences. The amphinomid *Hipponoe* was chosen to root the tree. Additional polychaete sequences were obtained from GenBank. Ambiguously aligned sections of 18S rDNA were excluded from the analysis. PAUP*4.0b10 (D. Swofford, Florida State University) was used for maximum parsimony analysis with characters equally weighted and gaps treated as missing data. Bootstrap values were estimated with the use of 1000 heuristic searches. Bayesian analysis used MrBayes v3.0B4 (J. Huelsenbeck, University of Rochester, NY) with partitions for stems and loops using RNA secondary structure prediction via GeneBee (L. Brodsky, Moscow State University, Russia). Likelihood models were chosen with the use of ModelTest 3.06 (D. Posada, Universidad de Vigo, Spain). Further details are available at (*4*).
23. Funding provided by the David and Lucile Packard Foundation, NSF (OCE9910799 and OCE0241613), and the South Australian Museum. Thanks to *Tiburon* pilots and *Western Flyer* crew for obtaining samples; J. Jones and R. Young for help with Bayesian analysis; R. Newbold and F. Pleijel for Latin advice; H. Schoppe, K. Rogers, and Adelaide Microscopy for microscopy support, and V. Orphan, L. Jahnke, T. Embaye, and K. Turk (NASA Ames Research Center) for nutritional characterization of the worms.

The Complete Genome Sequence of *Propionibacterium Acnes*, a Commensal of Human Skin

Holger Brüggemann,[1]*† **Anke Henne,**[1] **Frank Hoster,**[1] **Heiko Liesegang,**[1] **Arnim Wiezer,**[1] **Axel Strittmatter,**[1] **Sandra Hujer,**[2] **Peter Dürre,**[2] **Gerhard Gottschalk**[1]†

Propionibacterium acnes is a major inhabitant of adult human skin, where it resides within sebaceous follicles, usually as a harmless commensal although it has been implicated in acne vulgaris formation. The entire genome sequence of this Gram-positive bacterium encodes 2333 putative genes and revealed numerous gene products involved in degrading host molecules, including sialidases, neuraminidases, endoglycoceramidases, lipases, and pore-forming factors. Surface-associated and other immunogenic factors have been identified, which might be involved in triggering acne inflammation and other *P. acnes*–associated diseases.

The details of the involvement of *Propionibacterium acnes* in acne—the most common skin disease, affecting up to 80% of all adolescents in the United States—are still obscure. Several mechanisms have been proposed to account for its role in the disease (*1–5*). First, damage to host tissues and cells might be accomplished by bacterial enzymes with degradative properties, such as lipases (*2*). Second, immunogenic factors of *P. acnes* such as surface determinants or heat shock proteins (HSPs) might trigger inflammation (*4–6*). Other diseases are also associated with *P. acnes*, including corneal ulcers; endocarditis; sarcoidosis; cholesterol gallstones; allergic alveolitis; pulmonary angitis; and synovitis, acne, pustulosis, hyperostosis, and osteitis (SAPHO) syndrome (*7, 8*). Its genome sequence may provide a basis for finding alternative targets in therapy for acne and other *P. acnes*–associated diseases.

The genome of *P. acnes* strain KPA171202 (no. DSM 16379) consists of a single circular chromosome of 2,560,265 base pairs (*9, 10*) (supporting online text and fig. S1). We predicted and annotated 2333 putative genes. The sequenced strain exhibited 100% identity on the 16S ribosomal RNA level to several clinical *P. acnes* isolates, as well as to the well-studied laboratory strain P-37. The main features of the genome sequence and comparative analyses are described in the supporting online material (supporting online text and fig. S2).

[1]Göttingen Genomics Laboratory, Institute of Microbiology and Genetics, Georg-August-University Göttingen, Grisebachstraße 8, 37077 Göttingen, Germany. [2]Department of Microbiology and Biotechnology, University of Ulm, 89069 Ulm, Germany.

*Present address: Institut Pasteur, Laboratory of Genomics of Microbial Pathogens, 28 Rue du Dr. Roux, 75724 Paris, France.
†To whom correspondence should be addressed. E-mail: hbruegg@pasteur.fr (H.B.) and ggottsc@gwdg.de (G.G.)

The genome sequence offers insights into the traits that favor *P. acnes* as a ubiquitous commensal on human skin. Metabolic reconstruction reveals a capacity to cope with changing oxygen tensions, which confirms observations that strains of *P. acnes* can grow under microaerobic as well as anaerobic conditions. The genome sequence encodes all key components of oxidative phosphorylation that employs two terminal oxidases, a cytochrome aa_3 oxidase (PPA701/702) and a cytochrome d oxidase (PPA173-176), and a F_0F_1-type adenosine triphosphate synthase (PPA1238-1245). All genes of the Embden-Meyerhof pathway, the pentose phosphate pathway, and the tricarboxylic acid cycle are present. Under anaerobic conditions, strain KPA171202 can grow on several substrates such as glucose, ribose, fructose, mannitol, trehalose, mannose, *N*-acetylglucosamine, erythritol, and glycerol (*11*). In addition, several amino acid degrading pathways, similar to those of fermentative organisms, are present. Fermentative products are short-chain fatty acids, in particular propionic acid (*11*), whose production from pyruvate and methylmalonyl–coenzyme (CoA) is initialized by the methylmalonyl-CoA carboxyltransferase (PPA2005-2008). In addition to fermentative energy conservation, *P. acnes* possesses systems involved in anaerobic respiration such as nitrate reductase (PPA507-511), dimethyl sulfoxide reductase (PPA515-517), *sn*-glycerol-3-phosphate dehydrogenase (PPA2248-2250), and fumarate reductase (PPA950-952 and PPA1437-1439). Factors that might be involved in the life-style switch related to oxygen availability, as well as further aspects of the biology and biochemistry of *P. acnes*, are presented in the supporting online text.

Numerous genes have been found that can degrade and use host-derived substances (Table 1 and fig. S3). It has been proposed that free fatty acids, produced by *P. acnes* lipase activity on sebum, assist bacterial adherence and colonization of the sebaceous follicle (*2, 12*). In

addition to the previously identified gene for the secreted triacylglycerol lipase (PPA2105), various other lipase/esterase genes can be found in the genome. Three of these possess a C-terminal Leu-Pro-X-Thr-Gly (LPXTG)–type cell-wall sorting signal. As previously known, a hyaluronate lyase (PPA380) degrades hyaluronan, an important constituent of the extracellular matrix of connective tissues, and presumably aids bacterial invasion (13, 14). Genome sequencing revealed numerous additional enzymes putatively involved in host tissue degradation, such as two endoglycoceramidases (PPA644 and PPA2106) and four sialidases, two of them possessing a LPXTG motif (PPA1560 and PPA1569). Other degradative enzymes include a putative endo-β-N-acetylglucosaminidase (PPA990) and various extracellular peptidases. Five highly similar genes encoding homologs to CAMP (Christie, Atkins, Munch-Peterson) factors are also present in the genome of P. acnes (fig. S4). These factors are secreted proteins characteristic of some streptococcal species. CAMP factors have been shown to bind to immunoglobulins of the G and M classes and have long been known as pathogenic determinants. Recently, it was reported that CAMP factors can act as pore-forming toxins (15). These proteins could explain previously observed cytotoxic effects of P. acnes strains (16).

In several studies, the capability of P. acnes to interact with and stimulate the immune system has been investigated. Increased cellular, as well as humoral, immunity to P. acnes has been detected in patients with severe acne (3–5). The genome sequence encodes many factors with antigenic potential, for example, cell surface proteins that may also exhibit cell-adherent properties (Table 1 and fig. S3). Several of these (PPA1879-1881, PPA1955, PPA2127, and PPA2210) possess a C-terminal LPXTG-type cell-wall sorting signal that is required for attaching surface proteins to the cell-wall through the action of sortase (possibly encoded by candidate: PPA777), a mechanism employed by many Gram-positive bacteria (17). In all, 25 genes encoding proteins with a C-terminal LPXTG motif could be found in the genome. A few of these (PPA1880, PPA2127, and PPA2130) possess contiguous stretches of guanidine (G) or cytosine (C) residues, either in the putative promoter region or within the 5′ end. The sequences of several plasmids (of the shotgun library) that cover these regions were am-

Table 1. Selected factors of *P. acnes* that are putatively involved in degrading host molecules, conferring cell adhesion and/or mediating inflammation. ORF, open reading frame.

ORF number(s)	Function	Comment	Homology
380	Hyaluronate lyase	Polysaccharide lyase family protein	94% identity to *P. acnes* clone 49/51
570, 1035, 1101, 1224, 1425, 1631, 1745, 1761, 1839, 1953, 1967, 2036, 2142, 2150, etc.	Putative lipases/esterases	ORFs 570, 1745, and 2150 have a C-terminal LPXTG motif	Diverse
644, 2106	Endoglycoceramidases	Hydrolyzes glycosphingolipids	*Rhodococcus* sp., *Cyanea nozakii*
684	Sialidase L	Trans-sialidase	*Macrobdella decora* (leech)
685	Sialidase A	Exo-α-sialidase	*Clostridium perfringens*, *C. septicum*
687, 1198, 1231, 1340, 2108	CAMP factors	Immunoglobulin-binding, pore-forming toxin	*Streptococcus* species
990	Endo-β-N-acetylglucosaminidase		*Streptomyces plicatus*
1396	Putative hemolysin		*M. ulcerans, M. leprae*
1560, 1569	Sialidases/neuraminidases	With LPXTG motif	*Micromonospora viridifaciens*
1796, 2105	Triacylglycerol lipases	ORF 2105 is 100% identical to *gehA*	*P. acnes* P-37
1819–1821	Putative chitobiase/β-N-acetylhexosaminidase	Possible frameshift, with LPXTG motif	*Arthrobacter*
2139	Putative cutinase		*Botryotinia fuckeliana*
109	Myosin-crossreactive antigen		*Rhodopseudomonas palustris*
125–134, 145–150, 1692–1700, 1185, 1791, 2181	Glycosyltransferases, uridine diphosphate–N-acetylglucosamine 2-epimerase, polysaccharide biosynthesis proteins	Slime/capsular polysaccharide biosynthesis	*Staphylococcus aureus* and others
453, 1772, 1773	GroEL, Cpn10	The homolog of ORF 737 is a major immune reactive protein in mycobacteria	100% identity to GroEL and DnaK of *P. acnes* strain P-37
916, 2038	DnaJ2, DnaJ		
2039	GrpE		
2040, 1098	DnaK		
737	18-kD antigen		
721, 1962	Homologs of invasion-associated protein p60	NlpC/P60 family	*Listeria welshimeri*
571	Putative secreted surface protein	WD domain, senescence marker protein-30 domain	
765	Immunogenic protein, antigen 84	*divIVA* domain	*Corynebacterium efficiens, M. leprae, M. tuberculosis*
1983, 1984, 1906–1908, 1663–1666	Putative adhesions	Thrombospondin type 3 repeats, PKD and SEST domains	
1955	Surface-associated protein	RTX toxin domain, LPXTG motif	
1879, 1880, 1881, 2127	PTRPs	Putatively regulated by phase variation, LPXTG motifs	
1715, 2210, 2270	PTRPs	ORF 2210 has a LPXTG motif	
2130	Outer membrane protein A family protein	Putatively regulated by phase variation	

biguous with respect to the length of the poly(C)/(G) stretches (supporting online text and figs. S5 and S6). Such variable homopolymeric C or G stretches, generated in the course of replication because of slipped-strand mispairing, have been reported in other species to be involved in phase variation, an adaptation strategy to generate phenotypic variation (*18*) (supporting online text). Although showing no similarity to database entries, PPA1880 and PPA2127 both contain characteristic multiple repeats of the dipeptide proline-threonine [PT repetitive protein (PTRP)]. Such PT repeats have been detected in antigenic proteins of *Mycobacterium tuberculosis* (*19*). Additional PTRPs (PPA1715, PPA2210, and PPA2270) were found in *P. acnes* with characteristics of surface proteins, which may represent further host-interacting factors (Table 1 and supporting online text).

Several HSPs have been identified as major targets of the immune response in bacterial pathogens. Homologs of GroEL and DnaK were found and characterized in *P. acnes* (*6*). The genome sequence contains several more HSPs, such as DnaJ (PPA916 and PPA2038), GrpE (PPA2039), and an 18-kD protein (PPA737), the homologs of which in mycobacteria are major immune reactive proteins (*20*). Further proteins show similarities to known bacterial immunogenic factors, such as PPA765, to antigen 84 of *M. tuberculosis* and *M. leprae*, which is a highly immunogenic protein involved in the symptoms of multibacillary leprosy (*21*). Porphyrins are produced by *P. acnes* in high amounts, and these are also thought to be involved in inflammation (*1*) (supporting online text). In the presence of increasing oxygen tension, the interaction of molecular oxygen with released porphyrins generates toxic, reduced oxygen species, which can damage keratinocytes and lead to cytokine release (fig. S7).

In summary, the genome sequence clearly reveals many proteins involved in the ability of *P. acnes* to colonize and reside in human skin sites as well as a pronounced potential to survive a spectrum of environments. This capacity helps to explain the ubiquity of *P. acnes* and also its potential hazards, for example, the public health problems associated with Blood Bank contaminations, or the contamination of the human genome sequence database with *P. acnes* sequence. The GenBank entry AAH14236.1, a proposed human protein, is in fact a *P. acnes* protein (PPA1069).

References and Notes
1. K. T. Holland *et al.*, *Dermatology* **196**, 67 (1998).
2. J. E. Miskin, A. M. Farrell, W. J. Cunliffe, K. T. Holland, *Microbiology* **143**, 1745 (1997).
3. E. Ingham, *Curr. Opin. Infect. Dis.* **12**, 191 (1999).
4. A. Koreck, A. Pivarcsi, A. Dobozy, L. Kemeny, *Dermatology* **206**, 96 (2003).
5. U. Jappe, E. Ingham, J. Henwood, K. T. Holland, *Br. J. Dermatol.* **146**, 202 (2002).
6. M. D. Farrar, E. Ingham, K. T. Holland, *FEMS Microbiol. Lett.* **191**, 183 (2000).
7. E. Jakab *et al.*, *Yale J. Biol. Med.* **69**, 477 (1996).
8. T. Yamada *et al.*, *J. Pathol.* **198**, 541 (2002).
9. The sequence reported in this paper has been deposited in GenBank with accession no. AE017283.
10. Materials and methods are available as supporting material on *Science* Online.
11. S. Hujer, P. Dürre, unpublished data.
12. E. M. Gribbon, W. J. Cunliffe, K. T. Holland, *J. Gen. Microbiol.* **139**, 1745 (1993).
13. B. Steiner, S. Romero-Steiner, D. Cruce, R. George, *Can. J. Microbiol.* **43**, 315 (1997).
14. G. Makris, J. D. Wright, E. Ingham, K. T. Holland, *Microbiology* **150**, 2005 (2004).
15. S. Lang, M. Palmer, *J. Biol. Chem.* **278**, 38167 (2003).
16. Z. Csukas, B. Banizs, F. Rozgonyi, *Microb. Pathog.* **36**, 171 (2004).
17. D. Comfort, R. T. Clubb, *Infect. Immun.* **72**, 2710 (2004).
18. A. van der Ende *et al.*, *J. Bacteriol.* **177**, 2475 (1995).
19. K. K. Singh, X. Zhang, A. S. Patibandla, P. Chien, S. Laal, *Infect. Immun.* **69**, 4185 (2001).
20. R. J. Booth *et al.*, *Infect. Immun.* **61**, 1509 (1993).
21. P. W. Hermans *et al.*, *Infect. Immun.* **63**, 954 (1995).
22. This project was carried out within the framework of the Competence Network Göttingen "Genome Research on Bacteria" (GenoMik), financed by the German Federal Ministry of Education and Research (BMBF). A supporting grant was provided by the Niedersächsisches Ministerium für Wissenschaft und Kultur to the Göttingen Genomics Laboratory. H.B. is holder of a fellowship of the German Academy of Natural Scientists Leopoldina, funded by the BMBF.

Synthetic Mammalian Prions

Giuseppe Legname,[1,2]* Ilia V. Baskakov,[1,2]*† Hoang-Oanh B. Nguyen,[1] Detlev Riesner,[6] Fred E. Cohen,[1,3,4] Stephen J. DeArmond,[1,5] Stanley B. Prusiner[1,2,4]‡

Recombinant mouse prion protein (recMoPrP) produced in *Escherichia coli* was polymerized into amyloid fibrils that represent a subset of β sheet–rich structures. Fibrils consisting of recMoPrP(89–230) were inoculated intracerebrally into transgenic (Tg) mice expressing MoPrP(89–231). The mice developed neurologic dysfunction between 380 and 660 days after inoculation. Brain extracts showed protease-resistant PrP by Western blotting; these extracts transmitted disease to wild-type FVB mice and Tg mice overexpressing PrP, with incubation times of 150 and 90 days, respectively. Neuropathological findings suggest that a novel prion strain was created. Our results provide compelling evidence that prions are infectious proteins.

Prion diseases are responsible for some devastating neurological diseases, including Creutzfeldt-Jakob disease in humans, and present as infectious, genetic, and sporadic illnesses (*1*). The production of a new prion strain in Tg mice expressing an artificial, chimeric PrP transgene (*2*) encouraged us to renew our effort to produce synthetic wild-type prions. Earlier, we and others were unable to produce prion infectivity with the use of recombinant wild-type PrP refolded into β sheet–rich isoforms (*3*, *4*); hence, we turned to mutant PrPs.

Mice expressing high levels of MoPrP-(P101L), which harbors a mutation (Pro[101] → Leu) analogous to the human prion protein mutation that causes Gerstmann-Sträussler-Scheinker syndrome, develop neurodegeneration spontaneously at an early age. Brain extracts prepared from these Tg mice transmit prion disease to Tg mice expressing low levels of MoPrP(P101L), designated Tg196 mice (*5*). About 30% of Tg196 mice develop spontaneous illness at ~550 days of age. We synthesized a 55–amino acid peptide composed of MoPrP residues 89 to 143 with the P101L mutation and folded it into a β-rich conformation. The aggregated peptide produced neurologic dysfunction in Tg196 mice within ~1 year after inoculation, whereas the non–β-rich form did not (*6*). The incubation time for these mutant prions did not change upon serial passage (*7*). Even though the disease-causing MoPrP[Sc](P101L) readily formed amyloid in the brains of Tg196 mice, resistance of the protein to limited proteolysis could be demonstrated only under "mild" digestion conditions.

Although PrP amyloid deposition in brain is pathognomonic of prion disease, it is a nonobligatory feature (*8*). Moreover, full-length PrP[Sc] does not polymerize into amyloid, whereas N-terminally truncated PrP[Sc] (designated PrP 27-30) assembles into amyloid fibrils (*9*). On the basis of the foregoing findings, we reasoned that

[1]Institute for Neurodegenerative Diseases, Departments of [2]Neurology, [3]Cellular and Molecular Pharmacology, [4]Biochemistry and Biophysics, and [5]Pathology, University of California, San Francisco, CA 94143, USA. [6]Institut für Physikalische Biologie, Heinrich-Heine-Universität, 40225 Düsseldorf, Germany.

*These two authors contributed equally to this work.
†Present address: Medical Biotechnology Center, University of Maryland Biotechnology Institute, Baltimore, MD 21201, USA.
‡To whom correspondence should be addressed. E-mail: stanley@itsa.ucsf.edu

PrP amyloid represents a limited subset of β-rich PrPs, one or more of which might be infectious. Two protocols were used to produce the fibrils from recMoPrP(89–230) expressed in *E. coli* (*10*) (fig. S1); one used monomeric recMoPrP(89–230) to produce the amyloid fibrils (denoted "unseeded"), and the other used some of the unseeded fibrils as a seed for the production of nascent fibrils (denoted "seeded") (*11*).

The seeded and unseeded amyloid fibrils composed of recMoPrP(89–230) were inoculated intracerebrally into Tg(MoPrP,△23–88)9949/*Prnp*$^{0/0}$ mice (henceforth Tg9949 mice) expressing MoPrP(89–231) at a level 16 times that of normal PrP^C in Syrian hamster (SHa) brain. In Tg9949 mice, we do not know whether the glycolipid anchor of MoPrP(89–231) is attached to Ser^{230} or Ser^{231}. The two amyloid preparations as well as control phosphate-buffered saline (PBS) were prepared and injected on the same day. If contamination had been a problem, then the PBS-inoculated mice should have developed disease (table S1) unless the amyloids were selectively tainted.

Mice receiving recMoPrP(89–230) amyloid developed neurologic disease between 380 and 660 days after inoculation (Fig. 1A) (table S1). In earlier studies, uninoculated Tg9949 mice lived for more than 500 days without any signs of neurologic dysfunction (*12*). In the study reported here, Tg9949 mice were healthy at ~670 days of age and failed to show any signs of disease at 620 days after inoculation with PBS or 525 days with SHa Sc237 prions (table S1).

The shortest incubation times for Tg9949 mice inoculated with seeded amyloid and unseeded amyloid were 382 days and 474 days, respectively. Western blot analysis of brain homogenates of these two mice revealed that the Tg9949 mouse inoculated with seeded amyloid had more protease-resistant PrP^{Sc} than the brain of the unseeded

amyloid-inoculated mouse (Fig. 1B). Whether the different incubation times and diverse biochemical profiles reflect higher levels of PrP^{Sc} in the seeded amyloid relative to the unseeded amyloid, or whether these findings signify the creation of two different prion strains, remains to be established. No protease-resistant PrP was found by Western blotting in the brain of a Tg9949 mouse inoculated with PBS and killed at 530 days after inoculation; by this time, five of seven Tg9949 mice inoculated with seeded recMoPrP(89–230) amyloid had succumbed to disease (Fig. 1B). Consistent with the view that the amyloid fibrils contain low levels of MoPrP^{Sc}(89–230), we were unable to detect this protein in fibril preparations by immunoblotting, and we observed incubation times of ~500 days on first passage followed by shortening to ~250 days on second passage (table S1). For all strains of prions studied to date, low titer inocula produce prolonged incubation times that shorten upon subsequent passage (*13*). On first passage, mice do not become ill until the prion titers in their brains reach maximal levels. Whether the amyloid fibrils protected the small amounts of PrP^{Sc} found within them or modified the retention of PrP^{Sc} in brain after inoculation remains to be determined.

Residual MoPrP^C(89–231) after proteinase K (PK) digestion of brain homogenate from a PBS-inoculated Tg9949 mouse reflects the high level of PrP^C expression (Fig. 1B, lane 5), a problem encountered with some other Tg mice. Fortunately, the level of residual MoPrP^C(89–231) after PK digestion is sufficiently low as to not present problems with the interpretation of the data.

Tg9949 mice inoculated with seeded amyloid exhibited extensive vacuolation with gliosis in the cerebellum, hippocampus, brainstem, and white matter (fig. S2A). The

distribution, density, and morphology of the vacuoles were different for unseeded and seeded amyloid, raising the possibility that they represent two different prion strains (fig. S2A). Vacuolation, astrocytic gliosis, and PrP^{Sc} accumulation were more widely dispersed in gray matter regions in the brains of mice inoculated with unseeded amyloid relative to seeded amyloid (Fig. 2, A to D). The neuroanatomic distributions of vacuoles associated with unseeded and seeded amyloid were different from those found with mouse RML prions (compare fig. S2A with fig. S2B). With unseeded amyloid, the majority of vacuoles measured 20 to 50 μm in diameter (Fig. 2A), whereas most vacuoles from RML prions were 10 to 30 μm in diameter (Fig. 2E). From the seeded amyloid inoculum, smaller (10 to 20 μm) and larger (20 to 50 μm) vacuoles were evenly represented (Fig. 2C). With both unseeded and seeded amyloid, PrP^{Sc} was deposited in gray matter as relatively large solitary masses 5 to 20 μm in diameter and at the perimeter of many vacuoles (Fig. 2, B and D). In contrast, RML-infected mice exhibited fine granular accumulations of PrP^{Sc} (Fig. 2F).

Prions in the brains of Tg9949 mice inoculated with seeded amyloid were designated MoSP1 (mouse synthetic prion strain 1). Serial transmission of MoSP1 prions from Tg9949 mice to wild-type FVB mice and to Tg(MoPrP-A)4053 mice (henceforth Tg4053 mice) gave mean incubation times of 154 and 90 days, respectively (Fig. 3) (table S1). Tg4053 mice express MoPrP-A at a level 8 times that of SHaPrP^C in hamster brain (*14*). Protease-resistant PrP^{Sc} was found in the brains of both wild-type FVB and Tg4053 mice inoculated with MoSP1 prions (Fig. 1C).

Well-defined PrP amyloid plaques as well as numerous, densely packed, finely granular PrP^{Sc} deposits were identified in the brains of
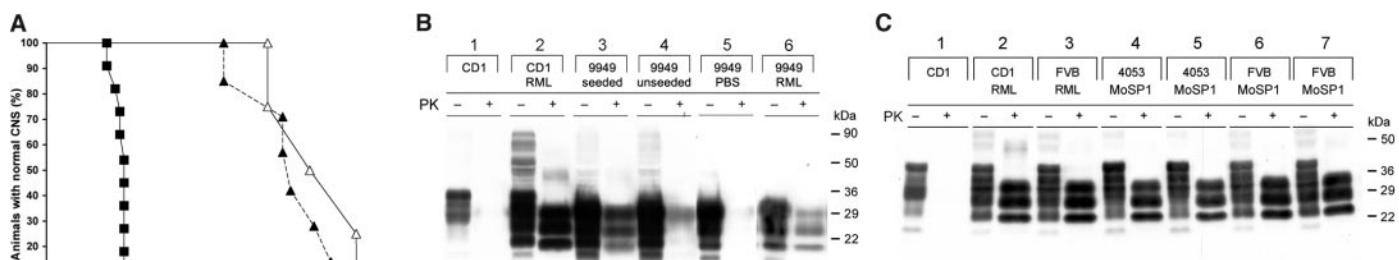

Fig. 1. Survival curves and immunoblots of Tg9949 mice. (**A**) Tg9949 mice inoculated with RML prions (■), seeded-amyloid recMoPrP(89–230) (▲), and unseeded-amyloid recMoPrP(89–230) (△). Amyloid fibrils were formed upon incubation of recMoPrP(89–230) (0.6 mg/ml) at 37°C in 3 M urea, 0.2 M NaCl, 50 mM sodium acetate buffer, pH 5.0, as described (*11*). The kinetics of fibril formation were monitored with a thioflavin T binding assay (*24*). Inocula were prepared by dialysis of the fibrils in PBS buffer, pH 7.2, for 2 days. Concentration of recMoPrP(89–230) in the inocula was ~0.5 mg/ml. (**B**) Immunoblot of PrP^{Sc} in brains of Tg9949 mice. Paired-samples lanes: 1, uninoculated normal CD1 mouse; 2, RML-inoculated CD1 mouse; 3, Tg9949 mouse inoculated with seeded amyloid; 4, Tg9949 mouse inoculated with unseeded amyloid; 5, Tg9949 mouse inoculated with PBS and killed at 580 days of age; 6, Tg9949 mouse inoculated with RML prions. (**C**) Immunoblot of PrP^{Sc} in brains of wild-type CD1, wild-type FVB, and Tg4053 mice. Paired-samples lanes: 1, uninoculated normal CD1 mouse; 2, RML-inoculated CD1 mouse; 3, RML-inoculated FVB mouse; 4 and 5, Tg4053 mouse inoculated with MoSP1 (brain homogenate from Tg9949 mice inoculated with seeded recMoPrP amyloid); 6 and 7, FVB mouse inoculated with MoSP1 (brain homogenate from Tg9949 mice inoculated with seeded recMoPrP amyloid). Symbols in (B) and (C): –, undigested control sample; +, samples subjected to limited PK digestion. Apparent molecular weights based on migration of protein standards are indicated.

both FVB (Fig. 4F) and Tg4053 mice inoculated with MoSP1 prions (*15*). The vacuolation scores (the percentage of an area occupied by vacuoles) were greater in the brains of Tg4053 mice than in those of FVB mice (fig. S3A). Additionally, the density of vacuoles (number per area) seen with MoSP1 prions was greater than that seen with RML prions in Tg4053 mice (fig. S3B). RML prions failed to cause vacuolation in the caudate nucleus, septal nuclei, and cerebellar white matter in Tg4053 mice. These findings argue that the MoSP1 strain remained stable during passage from Tg9949 to Tg4053 mice but seemed less stable upon passage in FVB mice, consistent with the pattern of PrP immunostaining (Fig. 4F).

The specificity of the response in Tg9949 mice to recMoPrP(89–230) polymerized into amyloid was established by the absence of disease after inoculation with PBS and Sc237 prions (table S1). When the healthy Tg9949 mice inoculated with Sc237 prions were killed at 525 days after inoculation, five of seven Tg9949 mice inoculated with the seeded amyloid were already ill.
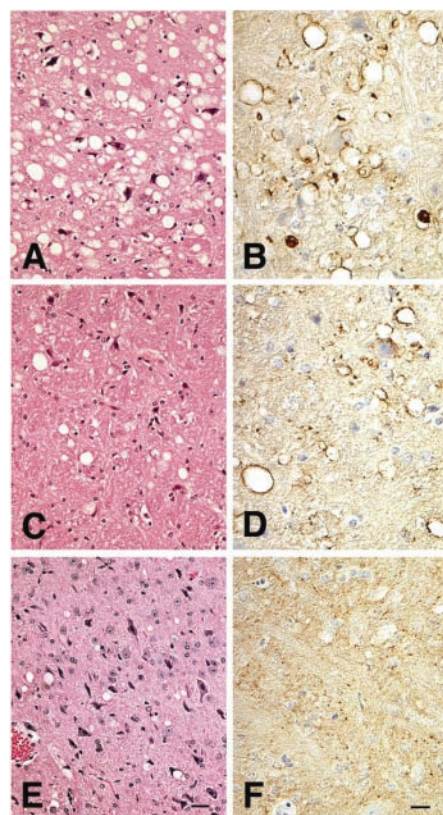
Many attempts to solubilize PrP^{Sc} under nondenaturing conditions and to renature PrP 27-30 have been unsuccessful (*16*, *17*); hence, progress in elucidating the structure of PrP^{Sc} has been slow (*18*). In purified fractions, PrP^{Sc} forms amorphous aggregates, whereas PrP 27-30 spontaneously assembles into amyloid polymers in the presence of detergent (*19*). Some investigators have argued that prions are infectious amyloids, ignoring findings that full-length PrP^{Sc} does not form amyloid (*20*, *21*). Notably, the fibrillar morphology of prion rods composed of PrP 27-30 disappeared after exposure to the organic solvent 1,1,1-trifluoro-2-propanol (TFIP) while prion infectivity was retained (*22*). Also of interest is a short peptide composed of 27 amino acids corresponding to PrP residues 100 to 126 that readily forms amyloid (*23*). On this background, we pursued investigations directed at producing synthetic prions using PrP amyloid as a surrogate marker for the folding of MoPrP(89–230) into a biologically active conformation. The rapidity and ease of measuring thioflavin T

binding that reflects amyloid formation (*24*) permitted us to determine conditions under which recMoPrP(89–230) would polymerize.

Our finding that amyloid fibrils harbor detectable levels of prion infectivity allows us to draw some conclusions about mammalian prions. First, PrP is both necessary and sufficient for infectivity. Second, neither the Asn-linked oligosaccharides nor the glycosylphosphatidylinositol anchor are required for prion infectivity, because recMoPrP(89–230) contains neither of these posttranslational modifications (*25–28*). Whether the Asn-linked oligosaccharides or the glycosylphosphatidylinositol anchor increase the efficiency of PrP^{Sc} formation for particular prion strains is unknown (*29*). Third, variations in PrP glycosylation are not required for prion diversity, and the biological information carried by distinct strains of prions resides in PrP^{Sc}, as previously suggested (*2*, *30*, *31*). Fourth, spontaneous formation of prions, which is thought to be responsible for sporadic forms of prion disease in livestock and humans, can occur in any mammal expressing PrP^C (*32*).
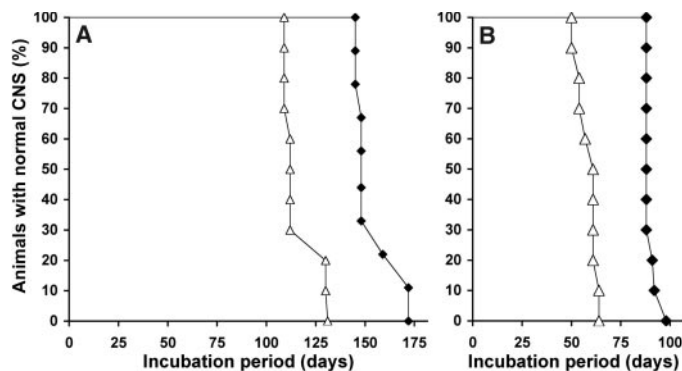

Fig. 2. The distinguishing neuropathological features of unseeded recPrP amyloid (**A** and **B**), seeded recPrP amyloid (**C** and **D**), and RML prions (**E** and **F**) in the pons of Tg9949 mice. Left column: hematoxylin and eosin (H&E) stain (scale bar, 50 μm); right column: immunohistochemistry of PrP^{Sc} by the hydrated autoclaving method using the PrP-specific HuM-R2 monoclonal antibody fragment (*43*) (scale bar, 25 μm).

Fig. 3. (**A**) Survival curves of FVB mice inoculated with RML (△) and MoSP1 (◆) prions. (**B**) Survival curves of Tg4053 mice inoculated with RML (△) and MoSP1 (◆) prions.
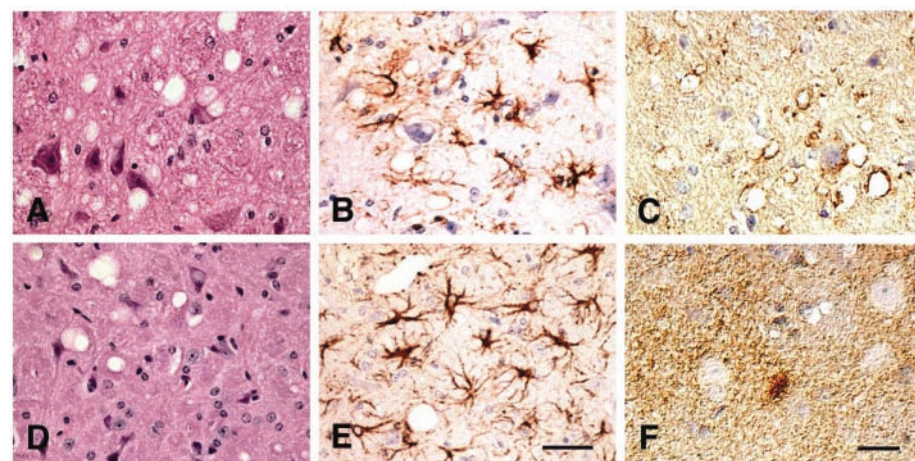

Fig. 4. Comparison of neuropathological changes in the pons associated with inoculation of seeded recPrP preparations into Tg9949 mice (**A** to **C**) and of MoSP1 (prions derived from clinically ill Tg9949 mice inoculated with seeded recPrP amyloid) inoculated into FVB mice (**D** to **F**). Both passages show the neurohistological characteristics of a prion disease: vacuoles (spongiform degeneration) shown by H&E staining (left); reactive astrocytic gliosis shown by glial fibrillary acidic protein immunohistochemistry (center); and accumulation of PrP^{Sc} visualized by hydrated autoclaving immunohistochemistry with HuM-R2 monoclonal antibody (right). Scale bars, 50 μm [(A), (B), (D), (E)], 25 μm [(C) and (F)].

Although we have used the polymerization of recMoPrP(89–230) into amyloid fibrils to generate prion infectivity, we hasten to add that other β-rich forms of recMoPrP(89–230) may also harbor infectivity. Preliminary results suggest that preparations of β-oligomers formed from recMoPrP(89–230) may also contain low levels of prion infectivity (*33*). Such findings emphasize the need to define optimal conditions for prion formation in vitro under which high levels of PrP^{Sc} can be generated. Moreover, previous difficulties in creating infectious prions in vitro from recPrPs enriched for β-structure may be due to the tendency of mammalian PrPs to fold into biologically irrelevant β-rich isoforms (*3*, *4*, *11*). In studies of fungal prions, the ease of assaying infectivity (*34*) and the ability to study millions of colonies made the creation of in vitro infectivity from recombinant proteins more tractable (*35–37*). Whereas yeast prions form within the cytoplasm (*38*), mammalian prions are thought to be produced on the cell surface in caveolae-like domains (*39*, *40*).

From Tg mouse studies, it is well established that templates improve the likelihood of forming an infectious β-rich isoform (*8*, *12*). In the studies reported here, we see evidence that seeded amyloid fibrils exhibit shorter incubation times than their unseeded progenitor (Fig. 1A). It remains to be determined whether this is due to the greater number of PrP^{Sc} molecules within seeded fibrils relative to unseeded fibrils, or whether this reflects strain differences.

Our results have important implications for human health. The formation of prions from recPrP demonstrates that PrP^{C} is sufficient for the spontaneous formation of prions; thus, no exogenous agent is required for prions to form in any mammal. Our findings provide an explanation for sporadic Creutzfeldt-Jakob disease for which the spontaneous formation of prions has been hypothesized. Understanding sporadic prion disease is particularly relevant to controlling the exposure of humans to bovine prions (*41*). That bovine prions are pathogenic for humans is well documented; more than 150 teenagers and young adults have already died from prion-tainted beef derived from cattle with bovine spongiform encephalopathy (*42*). Moreover, the sporadic forms of Alzheimer's and Parkinson's diseases as well as amyotrophic lateral sclerosis and the frontal temporal dementias are the most frequent forms of these age-dependent disorders, as is the case for the prion diseases. Important insights in the etiologic events that feature in these more common neurodegenerative disorders, all of which are caused by the aberrant processing of proteins in the nervous system, are likely to emerge as more is learned about the molecular pathogenesis of the sporadic prion diseases.

References and Notes

1. S. B. Prusiner, in *Prion Biology and Diseases,* S. B. Prusiner, Ed. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2004), pp. 89–141.
2. D. Peretz *et al., Neuron* **34**, 921 (2002).
3. I. Mehlhorn *et al., Biochemistry* **35**, 5528 (1996).
4. A. F. Hill, M. Antoniou, J. Collinge, *J. Gen. Virol.* **80**, 11 (1999).
5. K. K. Hsiao *et al., Proc. Natl. Acad. Sci. U.S.A.* **91**, 9126 (1994).
6. K. Kaneko *et al., J. Mol. Biol.* **295**, 997 (2000).
7. P. Tremblay *et al., J. Virol.* **78**, 2088 (2004).
8. S. B. Prusiner *et al., Cell* **63**, 673 (1990).
9. S. B. Prusiner *et al., Cell* **35**, 349 (1983).
10. See supporting data on *Science* Online.
11. I. V. Baskakov, G. Legname, M. A. Baldwin, S. B. Prusiner, F. E. Cohen, *J. Biol. Chem.* **277**, 21140 (2002).
12. S. Supattapone *et al., J. Virol.* **75**, 1408 (2001).
13. S. B. Prusiner, M. R. Scott, S. J. DeArmond, G. Carlson, in *Prion Biology and Diseases,* S. B. Prusiner, Ed. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2004), pp. 187–242.
14. G. C. Telling *et al., Genes Dev.* **10**, 1736 (1996).
15. G. Legname *et al.*, data not shown.
16. K. Post, D. R. Brown, M. Groschup, H. A. Kretzschmar, D. Riesner, *Arch. Virol. Suppl.*, 265 (2000).
17. H. Wille, S. B. Prusiner, *Biophys. J.* **76**, 1048 (1999).
18. C. Govaerts, H. Wille, S. B. Prusiner, F. E. Cohen, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 8342 (2004).
19. M. P. McKinley *et al., J. Virol.* **65**, 1340 (1991).
20. D. C. Gajdusek, *Br. Med. Bull.* **49**, 913 (1993).
21. J. T. Jarrett, P. T. Lansbury, Jr., *Cell* **73**, 1055 (1993).
22. H. Wille, G.-F. Zhang, M. A. Baldwin, F. E. Cohen, S. B. Prusiner, *J. Mol. Biol.* **259**, 608 (1996).
23. G. Forloni *et al., Nature* **362**, 543 (1993).
24. H. LeVine, *Protein Sci.* **2**, 404 (1993).
25. A. Taraboulos *et al., Proc. Natl. Acad. Sci. U.S.A.* **87**, 8262 (1990).
26. S. J. DeArmond *et al., Neuron* **19**, 1337 (1997).
27. P. Gambetti, P. Parchi, *N. Engl. J. Med.* **340**, 1675 (1999).
28. J. A. Mastrianni *et al., N. Engl. J. Med.* **340**, 1630 (1999).
29. J. Collinge, K. C. L. Sidle, J. Meads, J. Ironside, A. F. Hill, *Nature* **383**, 685 (1996).
30. R. A. Bessen, R. F. Marsh, *J. Virol.* **68**, 7859 (1994).
31. G. C. Telling *et al., Science* **274**, 2079 (1996).
32. S. B. Prusiner, *Annu. Rev. Microbiol.* **43**, 345 (1989).
33. G. Legname *et al.*, unpublished data.
34. R. B. Wickner, *Science* **264**, 566 (1994).
35. M. L. Maddelein, S. Dos Reis, S. Duvezin-Caubet, B. Coulary-Salin, S. J. Saupe, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 7402 (2002).
36. C. Y. King, R. Diaz-Avalos, *Nature* **428**, 319 (2004).
37. M. Tanaka, P. Chien, N. Naber, R. Cooke, J. S. Weissman, *Nature* **428**, 323 (2004).
38. M. M. Patino, J.-J. Liu, J. R. Glover, S. Lindquist, *Science* **273**, 622 (1996).
39. A. Gorodinsky, D. A. Harris, *J. Cell Biol.* **129**, 619 (1995).
40. P. J. Peters *et al., J. Cell Biol.* **162**, 703 (2003).
41. C. Casalone *et al., Proc. Natl. Acad. Sci. U.S.A.* **101**, 3065 (2004).
42. R. G. Will, M. P. Alpers, D. Dormont, L. B. Schonberger, in *Prion Biology and Diseases,* S. B. Prusiner, Ed. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2004), pp. 629–671.
43. D. Peretz *et al., Nature* **412**, 739 (2001).
44. We thank P. Culhane, K. Giles, P. Lessard, M. Scott, S. Supattapone, and the staff at the Hunters Point Animal Facility. Supported by NIH grants AG02132, AG10770, and AG021601, the G. Harold and Leila Y. Mathers Charitable Foundation, the Dana Foundation, and the Sherman Fairchild Foundation. G.L., D.R., F.E.C., S.J.D., and S.B.P. have a financial interest in InPro Biotechnology, which applies research on prions and prion diseases to promote human health.

Host-to-Parasite Gene Transfer in Flowering Plants: Phylogenetic Evidence from Malpighiales
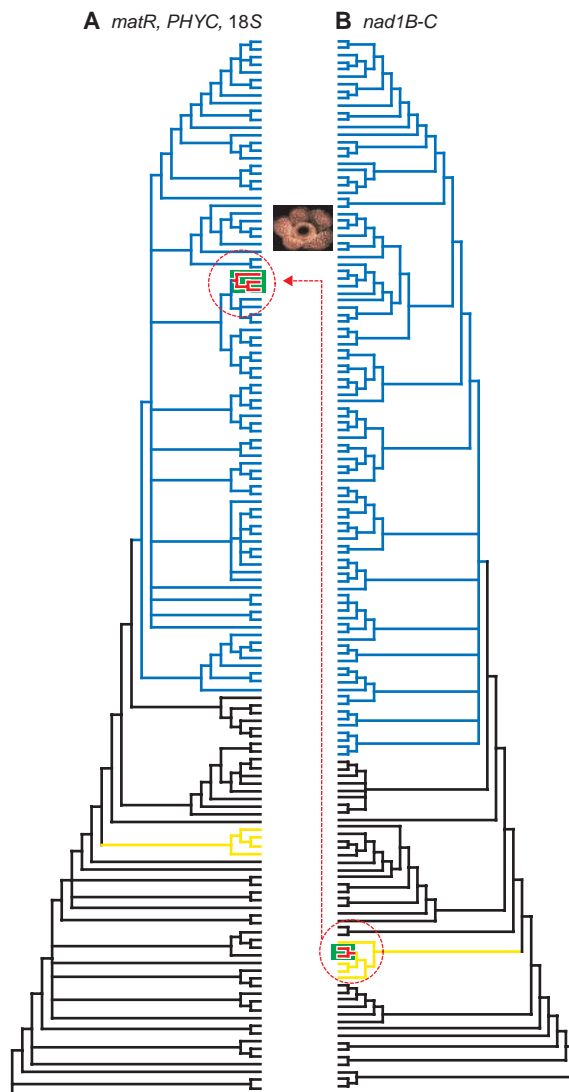
Charles C. Davis[1]* and Kenneth J. Wurdack[2]

Horizontal gene transfer (HGT) between sexually unrelated species has recently been documented for higher plants, but mechanistic explanations for HGTs have remained speculative. We show that a parasitic relationship may facilitate HGT between flowering plants. The endophytic parasites Rafflesiaceae are placed in the diverse order Malpighiales. Our multigene phylogenetic analyses of Malpighiales show that mitochrodrial (*matR*) and nuclear loci (18*S* ribosomal DNA and *PHYC*) place Rafflesiaceae in Malpighiales, perhaps near Ochnaceae/Clusiaceae. Mitochondrial *nad1B-C*, however, groups them within Vitaceae, near their obligate host *Tetrastigma*. These discordant phylogenetic hypotheses strongly suggest that part of the mitochondrial genome in Rafflesiaceae was acquired via HGT from their hosts.

Malpighiales are one of the most diverse clades of flowering plants uncovered in recent phylogenetic analyses. The order comprises 27 families (*1*) previously assigned to 13 different orders (*2*), including more than 16,000 species spanning tremendous morphological and ecological diversity (*3*). Recent surprising additions to Malpighiales are the endophytic holoparasites Rafflesiaceae (*4*), which lack leaves, stems, and roots, and rely entirely on their host plants, species of *Tetrastigma* (Vitaceae), for their nutrition. Despite their extreme vegetative reduction, they are unmistakable in flower, producing the largest flowers in the world, which mimic rotting flesh—an enticement to the carrion flies that pollinate them (*5*).

Barkman *et al.* (*4*) used mitochondrial (mt) *matR* sequences to place Rafflesiaceae firmly with Malpighiales [100%

Fig. 1. Two conflicting hypotheses about the phylogenetic placement of Rafflesiaceae. (**A**) The strict consensus of 136 angiosperms for combined mt *matR* and nuclear (*PHYC* and ribosomal 18*S*) data showing a well-supported (100% BP) Malpighiales clade (in blue), which includes all members of the order sensu APG II (*1*) plus Rafflesiaceae (in red; *Rafflesia*, *Rhizanthes*, and *Sapria*). (**B**) The strict consensus of 147 angiosperms for mt *nad1B-C* (the *nad1* intron 2 and part of the adjacent exons b and c) showing a well-supported (100% BP) Malpighiales clade, which includes all members of the order except Rafflesiaceae. Rafflesiaceae (*Rafflesia* and *Sapria*) are strongly placed (100% BP) in the basal eudicot family Vitaceae (in yellow) near their host genus, *Tetrastigma*. The dashed line is the hypothesized host/parasite HGT.

A *matR, PHYC,* 18*S*   B *nad1B-C*

Ochnaceae and Clusiaceae sensu lato, share tenuinucellate ovules (among mostly crassinucellate relatives) and staminal fusion with Rafflesiaceae (*2*, *3*).

The position of Rafflesiaceae based on *nad1B-C* provides a new example of horizontal gene transfer. If *nad1B-C* were vertically transmitted, as we believe to be the case for the other loci, we would expect Rafflesiaceae to group with Malpighiales. Instead, phylogenetic evidence from *nad1B-C* suggests that part of the mt genome in Rafflesiaceae originated from their hosts, *Tetrastigma* (either stem or crown group members), and was horizontally transferred to these obligate parasites. A similar horizontal gene transfer (HGT) of *nad1B-C* was recently reported (*11*) in seed plants, involving a transfer from an asterid to *Gnetum*. And Bergthorsson *et al.* (*12*) have documented several instances of mt HGT between distantly related angiosperm groups.

The underlying mechanism for HGT between sexually unrelated plants, however, has been elusive. Various pathogens have been suggested as primary vector agents (*11*, *12*). Our study documents a case in which there is no need to propose an intermediary vector for HGT. In these plants, the transfer appears to have been facilitated by the intimacy of the association between the host and the endophytic parasite, which lives its whole vegetative life as "an almost mycelial haustorial system," "ramifying and anastomosing throughout the [tissues of the] host" (*13*). This pattern may be an important mechanism by which parasites assemble their genetic architecture, and additional cases of HGT should be sought among other endophytic parasites and their hosts. It will also not be surprising if reciprocal genetic transfers are found to have occurred, from parasite to host.

References and Notes
1. The Angiosperm Phylogeny Group, *Bot. J. Linn. Soc.* **141**, 399 (2003).
2. A. Cronquist, *An Integrated System of Classification of Flowering Plants* (Columbia Univ. Press, New York, 1981).
3. P. F. Stevens, *Angiosperm Phylogeny Website*, ed. 4 (2001 onward) (www.mobot.org/MOBOT/research/Apweb).
4. T. J. Barkman, S.-K. Lim, K. Mat Salleh, J. Nais, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 787 (2004).
5. R. S. Beaman, P. J. Decker, J. H. Beaman, *Am. J. Bot.* **75**, 1148 (1988).
6. For primer and GenBank information see the supporting online material (SOM) on *Science* Online.
7. C. C. Davis, M. W. Chase, *Am. J. Bot.* **91**, 262 (2004).
8. For detailed information on materials and methods, data sets, phylogenetic analyses, and complete annotated figures see the SOM. We were unable to sample *PHYC* from all families of Malpighiales because it is probably absent from some clades (*7*). Hence, we analyzed these data in combination with 18*S* to ensure that all families were represented for the nuclear genome. We also examined the placement of Rafflesiaceae with 18*S* in two ways. First, we excluded the most divergent domains, V2 and V4, from the analysis (*14*) (322 of 1813 base pairs)] across all taxa. Second, we treated these domains as missing data

bootstrap percentage (BP)]. Their use of a single mt gene was appropriate in a family that has resisted placement with standard genetic loci. To further examine this placement, we obtained sequences representing all families of Malpighiales, all genera of Rafflesiaceae, and numerous basal eudicots for four loci from the mt and nuclear genomes (*6*). Low-copy nuclear genes are an underused resource for resolving the placement of problematic taxa, and *phytochrome C* (*PHYC*), as used here, has been useful for revealing relationships within Malpighiales (*7*).

Our phylogenetic analyses are summarized in Fig. 1 (*8*). The tree created from the *matR* and nuclear loci firmly (100% BP) place Rafflesiaceae within Malpighiales. In contrast, the mt locus *nad1B-C* suggests that Rafflesiaceae are not members of Malpighiales but belong (100% BP) in Vitaceae near their host *Tetrastigma*. Each of these mutually exclusive hypotheses cannot be attributable to contamination (*9*), and each receives strong support from parsimony analyses and from alternative topology tests.

Which of these conflicting hypotheses reflect the true species affinities of Rafflesiaceae? Vitaceae possess several synapomorphies that are rare among angiosperms, including sieve-tube plastids with starch and protein inclusions, pearl glands, stamens opposite the petals, and seeds with a cordlike raphe. If Rafflesiaceae were embedded in Vitaceae, as suggested by *nad1B-C*, we would expect species to possess at least some of these characters, but they do not (*2*, *3*). A definitive malpighialean sister group for Rafflesiaceae is unclear, given our data. However, the closest relatives suggested in the combined analysis (*10*),

¹Ecology and Evolutionary Biology, University of Michigan Herbarium, 3600 Varsity Drive, Ann Arbor, MI 48108–2287, USA. ²Department of Botany and Laboratories of Analytical Biology, Smithsonian Institution, 4210 Silver Hill Road, Suitland, MD 20746, USA.

*To whom correspondence should be addressed. E-mail: chdavis@umich.edu.

only for Rafflesiaceae. Both approaches yielded congruent topologies, with the latter (shown here) being much better resolved. The spurious placement of Rafflesiaceae as sister to all angiosperms when 18S is used (4) may be attributable to high divergence in these small domains and suggests that 18S may not be as useless for placing problematic taxa as previously suggested (4).

9. We took several precautions to avoid and detect contamination. First, our results were independently corroborated in the laboratories of each author. Second, Rafflesiaceae data were acquired before starting any work on Vitaceae. Third, if our DNA were cross-contaminated we would not expect such strongly conflicting results regarding the placement of Rafflesiaceae, given the same DNA. Nor would we expect such a high degree of sequence divergence of nad1B-C between Rafflesiaceae and Tetrastigma (or other Vitaceae we sampled). If contamination occurred, we would expect sequences to be nearly identical to those of other sampled Vitaceace, especially given the relatively low amount of sequence divergence between all accessions of Vitaceae in nad1B-C (0.51 to 0.95% sequence divergence). Not only is nad1B-C divergent within Rafflesiaceae (6.2%), but it is also highly divergent from other phylogenetically diverse Vitaceae (15) included here (2.5 to 3.3%).

10. Taxa in the combined analysis of 18S, PHYC, and matR were included if they were sampled for matR plus at least one of the two nuclear loci. We believe this analysis to represent the best estimate of Malpighiales phylogeny. Similar approaches combining multiple genes have provided powerful insights into angiosperm phylogeny where single gene studies have failed (16).

11. H. Won, S. S. Renner, Proc. Natl. Acad. Sci. U.S.A. 100, 10824 (2003).

12. U. Bergthorsson, K. L. Adams, B. Thomason, J. D. Palmer, Nature 424, 197 (2003).

13. J. Kuijt, The Biology of Parasitic Flowering Plants (Univ. of California Press, Berkeley, CA, 1969).

14. D. L. Nickrent, E. M. Starr, J. Mol. Evol. 39, 62 (1994).

15. M. J. Ingrouille et al., Bot. J. Linn. Soc. 138, 421 (2002).

16. Y.-L. Qiu et al., Nature 402, 404 (1999).

17. We thank W. Anderson, D. Boufford, Y.-L. Qiu, and P. Stevens. C.C.D. was supported by grants from the University of Michigan, NSF AToL 0431266, and the Michigan Society of Fellows. This paper is dedicated to V. Morrison.

KIF1A Alternately Uses Two Loops to Bind Microtubules

Ryo Nitta,[1] Masahide Kikkawa,[1,2] Yasushi Okada,[1] Nobutaka Hirokawa[1]*

The motor protein kinesin moves along microtubules, driven by adenosine triphosphate (ATP) hydrolysis. However, it remains unclear how kinesin converts the chemical energy into mechanical movement. We report crystal structures of monomeric kinesin KIF1A with three transition-state analogs: adenylyl imidodiphosphate (AMP-PNP), adenosine diphosphate (ADP)–vanadate, and ADP-AlFx (aluminofluoride complexes). These structures, together with known structures of the ADP-bound state and the adenylyl-(β,γ-methylene) diphosphate (AMP-PCP)–bound state, show that kinesin uses two microtubule-binding loops in an alternating manner to change its interaction with microtubules during the ATP hydrolysis cycle; loop L11 is extended in the AMP-PNP structure, whereas loop L12 is extended in the ADP structure. ADP-vanadate displays an intermediate structure in which a conformational change in two switch regions causes both loops to be raised from the microtubule, thus actively detaching kinesin.

To move along microtubules, kinesins (1) must alternate rapidly between tightly bound and detached states. How both dimeric (2, 3) and monomeric (4, 5) kinesins achieve this remains unclear. Because the binding energy in the strong-binding state [10 to 20 $k_B T$ (3, 4), where k_B is the Boltzmann constant and T is absolute temperature] is too large for rapid spontaneous release, the energy for fast detachment of kinesin from the microtubule must come from a step of the ATP hydrolysis cycle. Large change(s) in free energy are expected to occur during four steps: ATP binding, hydrolysis, phosphate release, and ADP release. Both conventional kinesin and KIF1A bind tightly to microtubules in the nucleotide-free state and in the ATP-bound state. In the ADP-bound state, conventional kinesin is detached from microtubules, whereas KIF1A is partially detached and diffuses freely along the microtubule. This is because loose binding of ADP-bound KIF1A is supported by the KIF1 family–specific K-loop at loop L12. A mutant KIF1A that lacks the K-loop detaches from the microtubule in the ADP-bound state, and the dissociation constant markedly varies depending on the type of bound nucleotide, as is true for conventional kinesin (4). For historical reasons, the tightly bound state is called the strong-binding state, and the fully or partially detached state is called the weak-binding state. Recent work detected the phosphate release from a mutant kinesin, which stalls before the detached state (6, 7). This means that detachment occurs just at or after the phosphate release. Thus, the active process to detach kinesin from the microtubule should occur at the transition from the strong-binding state to the weak-binding state.

The active detachment process can be detected in KIF1A because of its property of binding to the microtubule during adenosine triphosphatase (ATPase) cycling. The apparent dissociation constant of KIF1A in the presence of ATP is the weighted average of the equilibrium dissociation constant of various intermediate states during the ATPase turnover. Because the dissociation constant is not significantly different between two major intermediate states, the AMP-PNP–bound and ADP-bound states (Table 1) (fig. S1) (8), the apparent dissociation constant during the ATPase turnover was not expected to be fundamentally different from these values. However, the apparent dissociation constant in the presence of 2 mM ATP was twice the expected

[1]Department of Cell Biology and Anatomy, University of Tokyo, Graduate School of Medicine, Hongo, Bunkyo-ku, Tokyo 113-0033, Japan. [2]Department of Cell Biology, University of Texas Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, TX 75390, USA.

*To whom correspondence should be addressed. E-mail: hirokawa@m.u-tokyo.ac.jp

Table 1. (Apparent) equilibrium dissociation constants (K_d) for microtubules. K_d values are reported as means ± SEM of at least three independent experiments. Conditions: 2 mM nucleotide or its analog, 50 mM imidazole, 5 mM Mg-acetate, 1 mM EGTA, and 50 mM K-acetate, pH 7.4 at 27°C (nd, not determined).

Nucleotide	K_d (nM)			
	Wild type	L12†	L11‡	L8§
AMP-PNP	4.2 ± 1.3	6.0 ± 1.4	20.2 ± 4.0	25.0 ± 6.0
ADP	6.8 ± 2.5	23.5 ± 8.4	12.3 ± 4.0	26.5 ± 5.0
ATP*	10.8 ± 1.8	40.5 ± 11.8	nd	nd
ADP-AlFx	5.9 ± 1.5	7.1 ± 1.7	nd	nd
ADP-Vi	21.4 ± 4.3	167 ± 66	nd	nd

*ATP regeneration system was used to maintain ATP/ADP level. †L12: CK1 (4). ‡L11: K261A/R264A/K266A. §L8: K161A/R167A/R169A/K183A.

value (Table 1) (fig. S1). Considering that the duration of the intermediate states other than the AMP-PNP–bound and ADP-bound states is much shorter, this large apparent dissociation constant implies that KIF1A has another intermediate state with a considerably lower affinity to the microtubule, which likely corresponds to the actively detaching state.

Because this detached state must occur between the ATP- and ADP-bound states, we examined the KIF1A-microtubule interaction in the presence of two ADP-phosphate analogs, ADP-vanadate (ADP-Vi) and ADP-AlFx (9). Interestingly, the two analogs had different effects. ADP-Vi markedly lowered the affinity of KIF1A to the microtubule, whereas ADP-AlFx did not significantly change the affinity (Table 1) (fig. S1). Thus, ADP-Vi and ADP-AlFx stop the ATPase cycle of KIF1A at two different substates of the ADP-phosphate state. The ADP-Vi state likely corresponds to the actively detaching state (10). Extremely weak binding in the presence of ADP-Vi is not specific to KIF1A. For example, conventional kinesin and the *Drosophila* kinesin-like proteins ncd and nod also show extremely low affinity for microtubules in the presence of ADP-Vi, whereas ADP-AlFx increases affinity (11, 12). Because the affinity of conventional kinesin in the ADP-

AlFx state is closer to the affinity of the ATP-bound state than to the ADP-Vi or the ADP states (13), ADP-AlFx likely stops the ATPase cycle of kinesin at a state between the ATP-bound and ADP-Vi states (Fig. 1).

Thus, biochemical analyses suggest that the affinity of kinesin-type motors to the microtubules undergoes a drastic change during the hydrolysis of ATP into ADP. In the ATP-bound state and in the early ADP-phosphate state (ADP-AlFx), kinesin has its highest affinity to the microtubules (the strong-binding state). This state is followed by a transient late ADP-phosphate state (ADP-Vi), at which point kinesin has its lowest affinity (the actively detaching state). After phosphate release, kinesin in the ADP state has a moderate affinity (the weak-binding state).

To understand the atomic details of these mechanical states, we solved the crystal structures of KIF1A in the ADP-phosphate (ADP-AlFx and ADP-Vi) states and compared them with known structures of the ADP-bound state (1I5S) (14) and the AMP-PCP–bound state (1I6I) (14), as well as with an AMP-PNP–bound form solved here. As reported for myosin (15), AMP-PCP blocks the ATPase cycle before an isomerization step, which precedes ATP hydrolysis. Thus, the AMP-PCP–bound form would represent

a collision complex or a preisomerization state, rather than a prehydrolysis state, as suggested from the lack of interactions between the γ-phosphate of AMP-PCP and the two highly conserved residues in the switch regions (Ser^{215} and Gly^{251}) (see below) (fig. S2). In this study, we solved the structure in the AMP-PNP–bound state in order to examine the prehydrolysis state.

The x-ray crystal structures of the motor domain of KIF1A with ADP-Vi, ADP-AlFx, and AMP-PNP were refined to values of 2.2 Å, 2.6 Å, and 1.9 Å, respectively (see table S1 for details of our crystallographic work). To identify the regions that change conformation during ATP hydrolysis, we superimposed these three structures and the previously solved ADP-bound structure with the use of the Cα atoms of the conserved nucleotide-binding loop (P-loop: amino acids 97 to 104), which remains unchanged throughout ATP hydrolysis (Fig. 2) (14). The overall architectures of the ADP-Vi, ADP-AlFx, and AMP-PNP forms are similar to the previously solved ADP and AMP-PCP forms. However, although the central β sheets and surrounding α helices hardly change their conformations [root mean square deviation (RMSD) < 0.6 Å], three loop regions do undergo marked changes (16). These are the switch I region (loop L9; amino acids 202 to 218), the switch II cluster (loop L11, helix α4, loop L12, helix α5, and loop L13; amino acids 248 to 324), and the neck-linker region (strand β9; amino acids 353 to 361), which are all thought to play a crucial role in the nucleotide-driven motility of kinesin-type motors.

The switch I loop L9 changes its conformation, along with the movement and release of the γ-phosphate, during ATP hydrolysis. Before isomerization, loop L9 has a tight β-hairpin structure (AMP-PCP form, Fig. 3A). This structure melts so that, in the prehydrolysis state, the conserved serine residue (Ser^{215}) at the C-terminal end of the loop can approach the γ-phosphate (AMP-PNP form) (Fig. 3B) (fig. S2). After hydrolysis, the γ-phosphate moves toward the gap between the switch I loop L9 and the switch II loop L11 (ADP-AlFx form) (Fig. 3C) (fig. S3) (17), and the loop-to-helix transition of loop L9 likely supports the release of the γ-phosphate out of the nucleotide-binding pocket (ADP-Vi form) (Fig. 3D) (fig. S4) (18, 19).

This dynamic structural change in the switch I loop L9 might explain the rapid hydrolysis and release of the γ-phosphate. The γ-phosphate is efficiently released through the "back door" formed by the highly conserved residues Arg^{216} and Glu^{253} (fig. S4). In myosin, the salt bridge between these two residues in the prehydrolysis structure

Fig. 1. Kinesin (KIF1A)–microtubule hydrolysis cycle. Abbreviations: K, kinesin; M, microtubule; T, ATP; D, ADP; P, phosphate. Asterisks refer to the second ATP state (K*T) or the second ADP-phosphate state (K*DP) of kinesin. The main reaction pathway is enclosed (45).
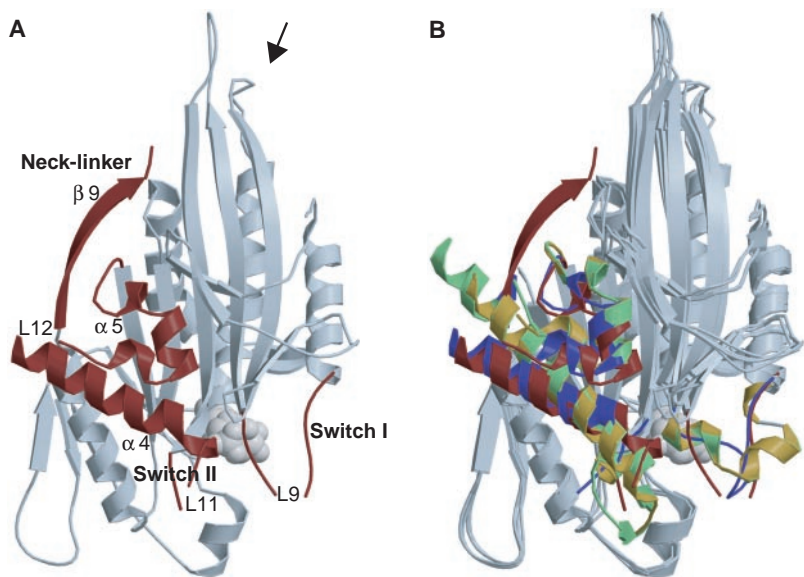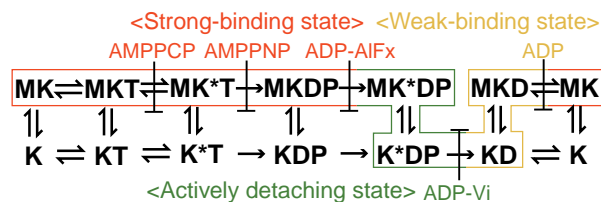

Fig. 2. Crystal structures of KIF1A. (**A**) The AMP-PNP form of KIF1A. The switch I, switch II, and neck-linker regions are highlighted in red. (**B**) Superposition of AMP-PNP, ADP-AlFx, ADP-Vi, and ADP structures. The AMP-PNP, ADP-AlFx, ADP-Vi, and ADP forms are shown in red, blue, green, and yellow, respectively.

closes this back door, which aligns the nucleophilic water with the γ-phosphate, and opening this back door generates a tunnel through which the γ-phosphate leaves the nucleotide-binding pocket (20, 21). KIF1A resembles myosin in the efficient release of the phosphate, which is generated by the swinging movement of the back door. The mutation of the back-door residue Arg[216] significantly slows ATP hydrolysis, which also supports the functional importance of these back-door residues in fast ATP hydrolysis (22–25).

The conformational change of the back-door loop L9 not only contributes to the rapid ATPase turnover, but also triggers large conformational changes in the microtubule-binding loops in the switch II cluster. Before hydrolysis, the conserved Gly[251] at the N-terminal end of the switch II cluster forms a hydrogen bond with the γ-phosphate (AMP-PNP form) (Fig. 3B) (fig. S2). The hydrolysis of ATP breaks this hydrogen bond and releases L11, the N-terminal loop of the switch II cluster (ADP-AlFx form) (Fig. 3C). The release of L11 triggers the rotation of the whole switch II cluster and the elongation of helix α4 from both ends by shortening and raising of both loops L11 and L12 during the phos-

phate release (Fig. 3, C and D). The raised conformation of loop L11 is stabilized by a salt bridge that forms between the conserved Glu[267] in L11 and the conserved Arg[216] in L9, and the hydrogen-bond network involving the γ-phosphate that connects switch I loop L9 and switch II loop L11 (fig. S4) (26, 27). Release of the phosphate breaks these bonds, which results in partial melting of the raised loops L11 and L12 in the ADP form. The rotation of the switch II cluster is supported by the salt bridge between the conserved Glu[170] in loop L8 and the conserved Arg[316] in helix α5 (Fig. 3, E and F) (28). This

Fig. 3. The conformational changes that occur in the two switch regions and the neck-linker region during ATP hydrolysis. (**A** to **D**) AMP-PCP (A), AMP-PNP (B), ADP-AlFx (C), and ADP-Vi (D) forms are shown in light brown, red, blue, and green, respectively. These panels are seen from the upper right in Fig. 2A (indicated by the black arrow). Nucleotides and the coordinating residues around them are shown as ball-and-stick models. Missing C-terminal residues and loops are shown by dashed lines. The structural details around the nucleotide-binding pocket are also shown in figs. S2 to S4 (46). (**E** and **F**) The conserved salt bridge between Glu[170] and Arg[316] (enclosed by a red circle), shown in ball-and-stick models in the AMP-PNP (E) and ADP-Vi (F) forms. Loop L8 and the switch II cluster are shown in dark blue and dark red, respectively. Nucleotides are shown as space-filling models.
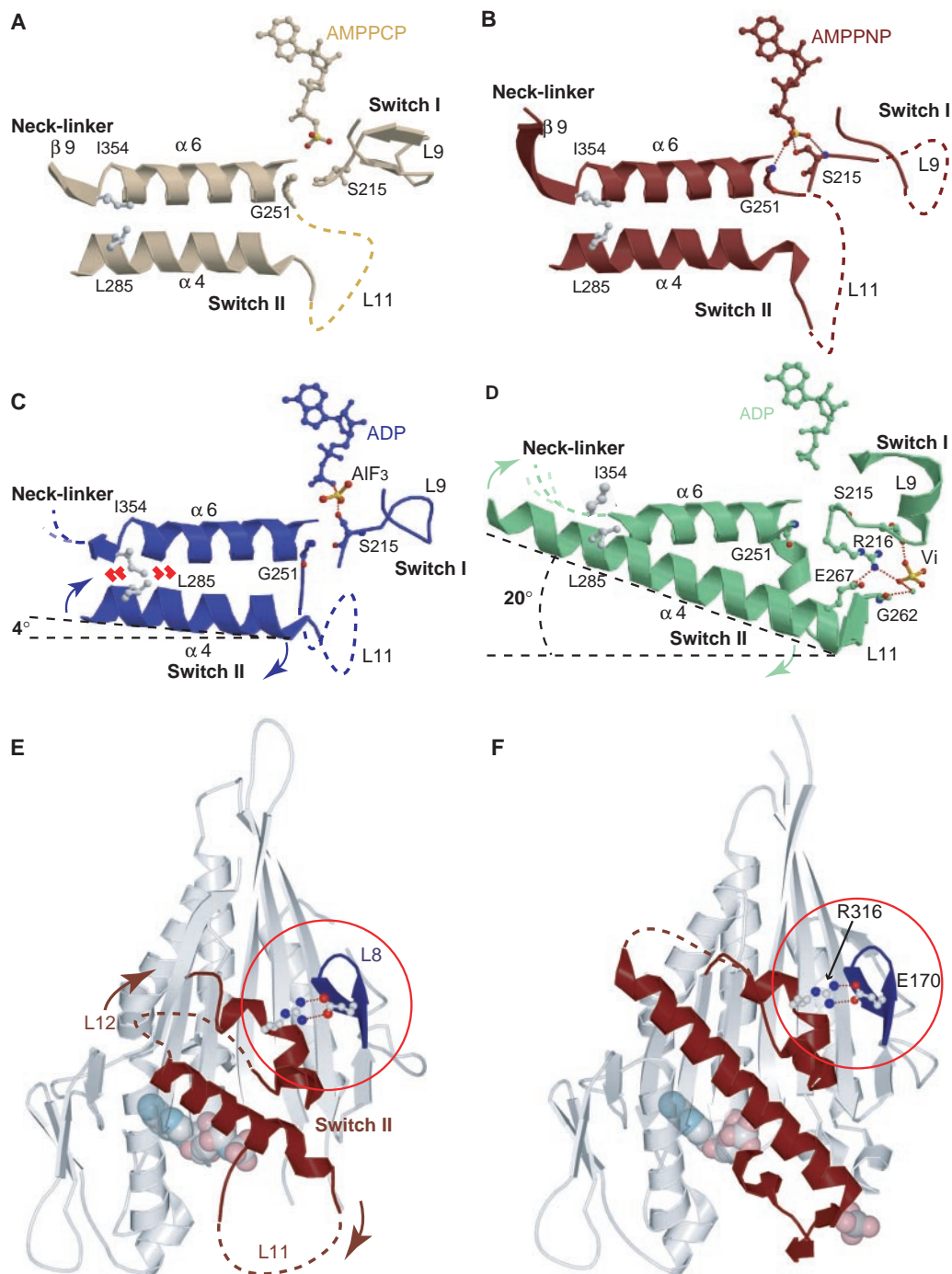
rotational movement also leads to the displacement of the neck, as we have proposed previously (*14*, *29*).

Thus, apparently, the rotational movement of the switch II cluster is supported by two conserved salt bridges: that between Arg[216] and Glu[253] (back door), and that between Glu[170] and Arg[316]. The back door might serve as a "latch" to keep the switch II cluster to the strained "up" position. The release of the latch triggers the rotation of this helix, which is pulled to the relaxed "down" position by the salt bridge between Glu[170] and Arg[316]. The interaction between loop L8 and the switch II cluster, mediated by the salt bridge between Glu[170] and Arg[316], might serve as the mechanochemistry coupler that couples the ATPase chemical reaction to the conformational change of the microtubule-binding surface. Consistent with this hypothesis, mutation of Glu[170] results in the decoupling of the ATPase turnover from the detachment from the microtubule (*6*). Interestingly, a recent structural study with another kinesin KIF2C suggested that L8 might serve as the sensor for the microtubule, triggering the closure of the back door (*30*). The conserved salt bridge between L8 and α5 might also contribute to closing the back door, because the mutation of either Glu[170] or Arg[316] considerably lowers the microtubule-activated ATPase activity (*31*).

Loop L11, helix α4, and loop L12 are the main microtubule-binding elements of kinesin (*4*, *31–33*); thus, their structural changes alter the affinity of kinesin to the microtubules. To examine the effect on microtubule binding, we docked the atomic structures of KIF1A and tubulin (*34*) to a KIF1A-microtubule complex whose structure was determined by cryogenic transmission electron microscopy (Fig. 4B) (*14*, *33*). In the AMP-PNP form, loop L11 of KIF1A is extended down to the H11′ helix of tubulin, and helix α4 of KIF1A is inserted deep into the intradimer groove of tubulin. KIF1A is tilted toward L11 so that L12 is extended up and away from the microtubule. During the hydrolysis of ATP, both loops L11 and L12 are raised from the microtubule, and only helix α4 supports the interaction between KIF1A and the microtubule (ADP-Vi form). The rotation of the
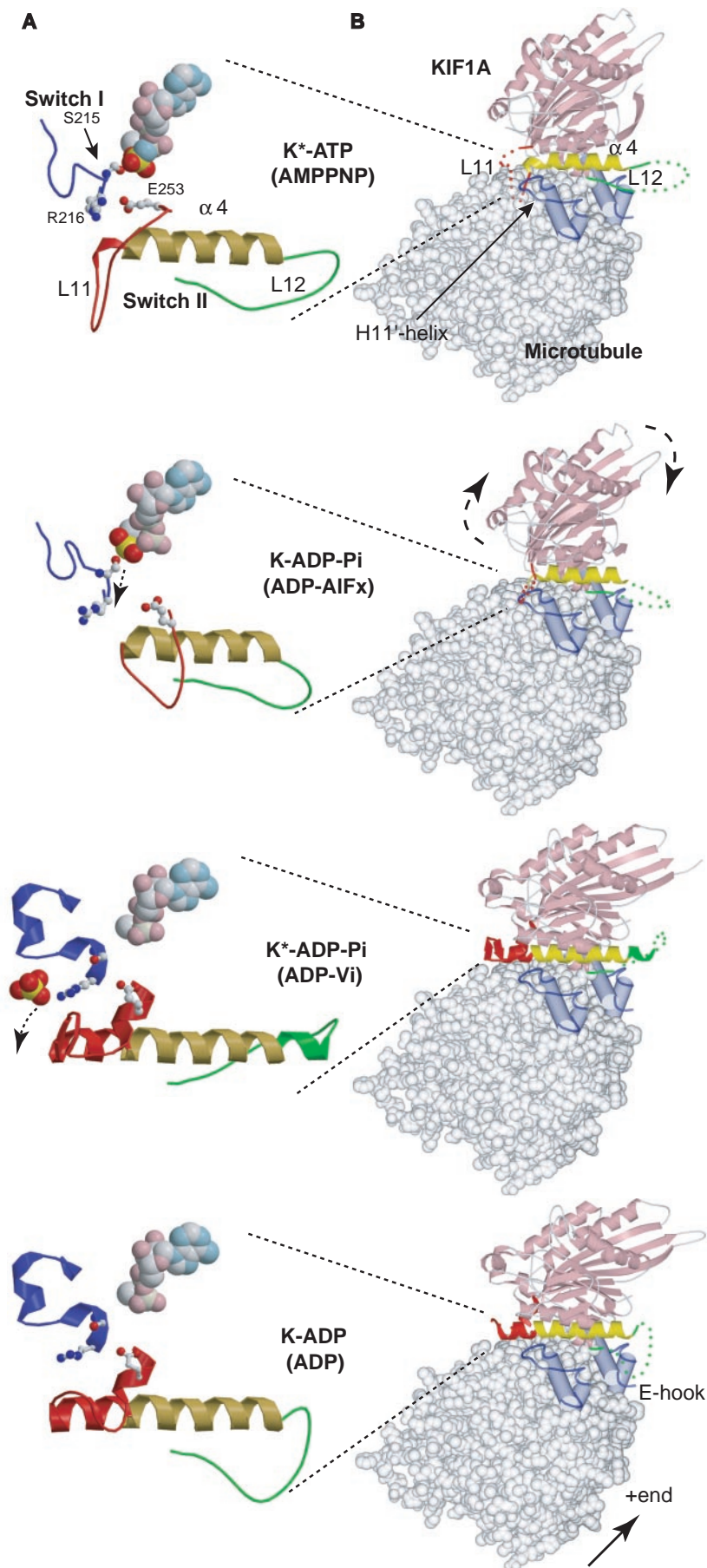

Fig. 4. Hydrolysis-induced conformational changes on the microtubule-binding surface of KIF1A. (**A**) Schematic illustration of the conformational changes in KIF1A. The switch I region (blue) and the switch II cluster (amino acids 251 to 269 in red, 270 to 288 in yellow, and 289 to 305 in green) are shown as ribbon models. Nucleotides are shown as space-filling models. The residues important for γ-phosphate release (Ser[215], Arg[216], and Glu[253]) are shown as ball-and-stick models. (**B**) The KIF1A-microtubule complex seen from the minus end of the microtubule. The microtubules (gray) are shown as space-filling models, except for helices H11 and H12 and the E-hook (blue). See movies S1 and S2 for details.

KIF1A core around helix α4 and the melting of loop L12 into the flexible conformation then allow loop L12 to interact with the C-terminal E-hook of tubulin in the ADP form while loop L11 is still raised.

These structural models suggest a third transient mechanical state, in which kinesin detaches itself from the microtubule, between the strong-binding and weak-binding states (Fig. 4B). In the ATP-bound state, kinesin is tightly fixed to the microtubule by interactions between loop L11 of kinesin and the H11′ helix of tubulin (the strong-binding state). ATP hydrolysis then generates the actively detaching state by raising the two main microtubule-binding loops L11 and L12. Thus, kinesin is only weakly supported by helix α4 and other minor binding elements. After phosphate release, the flexible loop L12 is extended down onto the tubulin C-terminal E-hook, and this flexible interaction allows one-dimensional diffusion in the weak-binding state (4, 5, 32). In short, three successive states of the kinesin-microtubule interaction exist during ATP hydrolysis, and kinesin uses the energy of ATP hydrolysis to detach loop L11 from the microtubule and to extend another loop (L12) to the microtubule.

We tested this proposal of alternating use of the two microtubule-binding loops by introducing mutations to these loops. The positively charged residues Lys[261], Arg[264], and Lys[266] in loop L11 were mutated to Ala, and this L11 mutant (K261A/R264A/K266A) was compared with an L12 mutant (CK1) (4) and an L8 mutant in which Lys[161], Arg[167], Arg[169], and Lys[183] were mutated to Ala (K161A/R167A/R169A/K183A). Loop L8 was selected for mutation because it corresponds to the third microtubule-binding region and does not change its conformation or position relative to the microtubule during ATPase turnover (14, 33). Relative to the wild-type protein, all of the mutants showed similar ATPase turnover rates but a lower affinity to the microtubule. As previously shown (4), the mutation in loop L12 selectively weakened affinity in the ADP state but not in the AMP-PNP state (Table 1). By contrast, the mutation in loop L11 had a stronger effect in the AMP-PNP state than in the ADP state. The mutation in loop L8 affected both forms to the same extent. Thus, the mutational studies support the proposal that KIF1A uses two loops, L11 and L12, in an alternating manner during ATPase turnover.

This alternating use of microtubule-binding loops provides structural support for our model of KIF1A's motility (4, 5, 32, 35). The model assumes that KIF1A alternates between two different binding states: the rigor-binding state with the affinity biased toward the forward tubulin subunit, and the diffusive binding state that allows one-dimensional diffusion. The rigor-binding loop L11 is extended toward the forward

tubulin subunit. This configuration of loop L11 gives directional bias, or asymmetric binding potential, to the rigor-binding state. The force and binding energy supported by this binding are more than 10 pN and 10 $k_B T$ (3, 4), respectively. This strong binding is broken by the energy of ATP hydrolysis. ATP hydrolysis causes conformational changes that first raise loop L11 from the microtubule and then bring loop L12 into contact with the microtubule. The flexible tether between loop L12 and the E-hook of tubulin allows diffusive binding in the ADP form. Thus, KIF1A kinesin takes three structures successively: L11 down, both L11 and L12 up, and L12 down. It is therefore possible to surmise that a fourth structure might exist: both L11 and L12 down (at the transition from the L12 down, ADP state, to the L11 down, ATP state). Bringing both L11 and L12 to the microtubule might open the nucleotide-binding pocket and accelerate ADP release. Further structural analysis of the nucleotide-free state or the ADP-release intermediate state of KIF1A will be required to examine this possibility.

Finally, we note that the design principle of KIF1A is similar to that of heterotrimeric guanine nucleotide–binding proteins (G proteins) (36, 37) and other proteins such as protein kinases (38). The binding energy of KIF1A to the microtubule is used to achieve a mechanical step of KIF1A (5). With G proteins and protein kinases, the binding energy to an effector molecule or substrate is used to cause structural changes. Nucleophilic attack by water molecules or serine residues on the γ-phosphate catalyzes hydrolysis. Hydrolysis triggers a conformational change at the binding surface (the switch II helix) that detaches the enzyme from the target molecule. The binding between a G protein and its effector molecule, or between kinase and its substrate, is so strong that energy is required for dissociation. The energy of hydrolysis is used for this active detachment, which is necessary to cycle the reaction. This conserved strategy might reflect the evolutionary pathway of these classes of proteins, and also suggests a design principle for nanomachines.

References and Notes

1. N. Hirokawa, *Science* **279**, 519 (1998).
2. M. Nishiyama, E. Muto, Y. Inoue, T. Yanagida, H. Higuchi, *Nature Cell Biol.* **3**, 425 (2001).
3. M. J. Schnitzer, K. Visscher, S. M. Block, *Nature Cell Biol.* **2**, 718 (2000).
4. Y. Okada, N. Hirokawa, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 640 (2000).
5. Y. Okada, H. Higuchi, N. Hirokawa, *Nature* **424**, 574 (2003).
6. L. M. Klumpp, A. Hoenger, S. P. Gilbert, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 3444 (2004).
7. I. M. Crevel *et al.*, *EMBO J.* **23**, 23 (2004).
8. Although the biochemically measured apparent dissociation constant is not significantly different among the AMP-PNP–bound, nucleotide-free, and

ADP-bound states, each kinesin molecule behaves differently on the microtubule in the ADP-bound state (one-dimensional diffusion) relative to the other two states (rigor-binding) (4, 5). Biochemical analysis using [α-^{32}P]ADP indicates that more than 90% of the active sites of kinesin molecules are occupied with ADP. These results, as well as previous results with the L12 mutant CK1 (4), support that the result with excess ADP (2 to 10 mM) reflects the behavior of kinesin with ADP in the active site.
9. Our previous experiment with ADP plus phosphate failed to detect the actively detaching state (4). This might be due to the irreversibility of the phosphate release. As reported with myosin (39), our preliminary biochemical results and our current structural study (fig. S4) suggest that exogenously added phosphate apparently does not bind to KIF1A. Thus, in our previous experiments with ADP and phosphate, most of the KIF1A will be in the ADP-bound state, not in the ADP-phosphate state. That is why we used analogs, which are expected to better mimic the true intermediate states of ATPase turnover.
10. The increase in the value of the dissociation constant (K_d) of the ADP-Vi state relative to other states suggests that the dissociation of kinesin is accelerated by a factor of at least 5. Our preliminary results with single-molecule imaging suggest that the off-rate is faster in the ADP-Vi state than in other states by a factor of more than 300.
11. E. Pechatnikova, E. W. Taylor, *Biophys. J.* **77**, 1003 (1999).
12. H. J. Matthies, R. J. Baskin, R. S. Hawley, *Mol. Biol. Cell* **12**, 4000 (2001).
13. S. S. Rosenfeld, B. Rener, J. J. Correia, M. S. Mayo, H. C. Cheung, *J. Biol. Chem.* **271**, 9473 (1996).
14. M. Kikkawa *et al.*, *Nature* **411**, 439 (2001).
15. J. F. Koretz, E. W. Taylor, *J. Biol. Chem.* **250**, 6344 (1975).
16. The RMSDs of the switch I region (amino acids 202 to 218) are 2.0 Å (AMP-PNP versus ADP-AlFx forms), 4.5 Å (ADP-AlFx versus ADP-Vi forms), and 4.0 Å (ADP-Vi versus AMP-PNP forms). Those of the switch II region (amino acids 248 to 324) are 0.8 Å (AMP-PNP versus ADP-AlFx forms), 1.4 Å (ADP-AlFx versus ADP-Vi forms), and 2.1 Å (ADP-Vi versus AMP-PNP forms).
17. Although the γ-phosphate is expected to be nucleophilically attacked by water, by analogy to myosin (20, 40), the γ-phosphate in our transition structure with ADP-AlFx coordinates with Ser[215] in the switch I region. This coordination could possibly be an artifact due to the use of a nucleotide analog. However, the difference map of AlF$_3$ indicates that Al^{3+} exists in trigonal bipyramidal configuration with partial bonds to the β-phosphate and the hydroxyl group of Ser[215] (Fig. 3C) (fig. S3). This coordination suggests another possibility: that the γ-phosphate is nucleophilically attacked by conserved Ser[215], as occurs in protein kinases. Further studies will be necessary to clarify this issue.
18. The conserved residue Arg[216] in the switch I loop L9 has been suggested to be the important phosphate-binding site near the nucleotide-binding pocket (41).
19. The possibility that our structure of ADP-Vi form might only reflect a structural variation of ADP is excluded by the following results. First, ADP-bound KIF1A never takes an ADP-Vi like structure, even under conditions identical or similar to the crystallization condition for the ADP-Vi–bound KIF1A. Similarly, ADP-Vi–bound KIF1A never takes an ADP-like structure, even under the crystallization condition for the ADP form. This is very different from the case with the ADP and AMP-PCP forms (14). ADP-bound KIF1A can take an AMP-PCP–like structure (but not an AMP-PNP–like structure) under the crystallization condition for the AMP-PCP form. This reflects that the energy barrier for the transition between the ADP-Vi and ADP structures is much higher than that between the ADP and AMP-PCP structures. Second, only preformed ADP-Vi-KIF1A complex takes this ADP-Vi structure. We could not obtain a crystal of ADP-Vi form by soaking vanadate solution to ADP-KIF1A crystals or by the addition of vanadate to the mother liquid with which ADP-bound KIF1A is crystallized. These results strongly suggest that our ADP-Vi structure reflects a structure of a stable intermediate state within the ATPase cycle, rather than

another structural variant or an artificial vanadate adduct to ADP-KIF1A.

20. R. G. Yount, D. Lawson, I. Rayment, *Biophys. J.* **68**, 47S (1995).
21. M. Furch, S. F. Becker, M. A. Geeves, K. C. Holmes, D. J. Manstein, *J. Mol. Biol.* **290**, 797 (1999).
22. M. Yun, X. Zhang, C. G. Park, H. W. Park, S. A. Endow, *EMBO J.* **20**, 2611 (2001).
23. C. M. Farrell, A. T. Mackey, L. M. Klumpp, S. P. Gilbert, *J. Biol. Chem.* **277**, 17079 (2002).
24. Recently, a new model was proposed for myosin on the basis of a new structure (*42*). Our structures and model are not inconsistent with this new mechanism. As proposed for the new "trapdoor" mechanism, kinesin has two "doors" that open and close asynchronously. The door on the switch I side is opened and the door on the switch II side is closed ("O/C conformation") in the preisomerization state. This conformation might contribute to the holding of the entering ATP. Before hydrolysis, KIF1A takes a "C/C conformation." During hydrolysis, it transiently takes a "C/O conformation" for the effective nucleophilic attack. Then, it takes an "O/O conformation," which opens the route through which the phosphate is released. Further studies will be necessary to determine whether kinesin works by a back-door or a trapdoor mechanism.
25. Here, two differences between kinesins and myosins should be noted: (i) The nucleotide-binding pocket of myosin is enclosed, whereas that of kinesin is rather exposed to the surface; (ii) the back door in the prehydrolysis state of myosin is rigidly closed, whereas that of kinesin is only partially closed. These differences might affect the kinetics of kinesin ATPase much faster than those of myosins (*32*, *43*). Further structural study of the KIF1A-microtubule complex will be needed to clarify these differences.
26. The possibility that the vanadate moiety on Arg[216] only reflects an artificial coordination of the solvent ion in crystal can be excluded by the following reasons. First, the position of vanadate is analogous to that in myosin. With myosin, vanadate is found on the corresponding back-door residue through which the produced phosphate is thought to be released. Second, vanadate is located at the same or similar position even in solution. The position of vanadate in solution was determined by the photo cleavage experiment. With myosins, the cleavage site (P-loop) agrees well with its ADP-Vi structure (*40*). Similarly, KIF1A was cleaved at loop L11, as expected from our ADP-Vi structure. This confirms that our ADP-Vi structure actually reflects the solution structure.
27. This stabilization of loop L11 of the ADP-Vi form is also evident from the low temperature factors of the residues in loop L11 [mean *B* value of the main-chain atoms of loop L11 (amino acids 251 to 270), 39.5 Å², overall mean *B* value, 32.0 Å²]. In the ADP form, the residues from Ala[255] to Ala[260] in loop L11 are disordered because of its flexibility.
28. This arginine is conserved among kinesins, although some kinesins have arginine at the residue corresponding to 313, a position just next to 316 on helix α5. For example, in conventional kinesin (3KIN), Arg[313] forms a salt bridge to Glu[170].
29. Before ATP hydrolysis, the neck-linker β9 (amino acids 353 to 361) is fully docked [except for the (His)₆ tag at the C terminus] to the catalytic core (AMP-PCP and AMP-PNP forms) (Fig. 3, A and B). During hydrolysis, the rotational movement of the switch II helix α4 pushes the neck-linker β9 down through the hydrophobic interaction between the conserved Leu[285] of helix α4 and the conserved Ile[354] of strand β9 (ADP-AlFx form) (Fig. 3C). This leads to the complete detachment of the neck-linker from the catalytic core (ADP-Vi and ADP forms) (Fig. 3D).
30. T. Ogawa, R. Nitta, Y. Okada, N. Hirokawa, *Cell* **116**, 485 (2004).
31. G. Woehlke *et al.*, *Cell* **90**, 207 (1997).
32. Y. Okada, N. Hirokawa, *Science* **283**, 1152 (1999).
33. M. Kikkawa, Y. Okada, N. Hirokawa, *Cell* **100**, 241 (2000).
34. J. Lowe, H. Li, K. H. Downing, E. Nogales, *J. Mol. Biol.* **313**, 1045 (2001).
35. This active detachment mechanism is also applicable to dimeric kinesin, including conventional kine-

sin. Hydrolysis of one head triggers the conformational change of its switch II cluster, both the microtubule-binding loop raised and the switch II helix rotated. The rotation of this helix pushes the neck-linker by which two heads coordinate. Thus, our study provides structural evidence of the idea that the active detachment is coupled with the coordination of two heads (*44*). In other words, active detachment of one head precedes the hydrolysis of the second head, the second step.
36. S. R. Sprang, *Annu. Rev. Biochem.* **66**, 639 (1997).
37. J. J. Tesmer, S. R. Sprang, *Curr. Opin. Struct. Biol.* **8**, 713 (1998).
38. D. A. Johnson, P. Akamine, E. Radzio-Andzelm, M. Madhusudan, S. S. Taylor, *Chem. Rev.* **101**, 2243 (2001).
39. C. F. Midelfort, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 2067 (1981).
40. C. A. Smith, I. Rayment, *Biochemistry* **35**, 5404 (1996).
41. C. V. Sindelar *et al.*, *Nature Struct. Biol.* **9**, 844 (2002).
42. T. F. Reubold, S. Eschenburg, A. Becker, F. J. Kull, D. J. Manstein, *Nature Struct. Biol.* **10**, 826 (2003).
43. M. A. Geeves, *Biochem. J.* **274**, 1 (1991).
44. A. T. Mackey, S. P. Gilbert, *Biochemistry* **39**, 1346 (2000).
45. Both conventional kinesin and KIF1A share this pathway. The only major difference is the affinity of kinesin to microtubule in the ADP state (yellow). The K-loop of KIF1A increases its affinity in the ADP state (*4*), shifting the equilibrium to the associated state

(MKD). With K-loop mutant KIF1A as well as conventional kinesin, the equilibrium is shifted to the dissociated state (KD).
46. See supporting material on *Science* Online.
47. We thank D. R. Tomchick, N. Ota, S. Wakatsuki, N. Igarashi, H. Sawa, Y. Wakabayashi, and their colleagues for technical suggestions; H. Fukuda, H. Sato, and M. Sugaya for technical and secretarial assistance; and H. Miki, M. Kawagishi, H. Yajima, T. Ogawa, E. Nitta, and our colleagues for discussions and technical assistance. Supported by a Center of Excellence Grant-in-Aid from the Ministry of Education, Culture, Sports, Science and Technology of Japan (N.H.). Coordinates for the four new structures of KIF1A have been deposited in the PDB database and are available under the following accession codes: both 1VFV and 1VFW for the AMP-PNP form, 1VFX for the ADP-AlFx form, and 1VFZ for the ADP-Vi form.

Crystal Structures of Human Cytochrome P450 3A4 Bound to Metyrapone and Progesterone

Pamela A. Williams,[1] Jose Cosme,[1] Dijana Matak Vinković,[1] Alison Ward,[1]* Hayley C. Angove,[1] Philip J. Day,[1] Clemens Vonrhein,[2] Ian J. Tickle,[1] Harren Jhoti[1]†

Cytochromes P450 (P450s) metabolize a wide range of endogenous compounds and xenobiotics, such as pollutants, environmental compounds, and drug molecules. The microsomal, membrane-associated, P450 isoforms CYP3A4, CYP2D6, CYP2C9, CYP2C19, CYP2E1, and CYP1A2 are responsible for the oxidative metabolism of more than 90% of marketed drugs. Cytochrome P450 3A4 (CYP3A4) metabolizes more drug molecules than all other isoforms combined. Here we report three crystal structures of CYP3A4: unliganded, bound to the inhibitor metyrapone, and bound to the substrate progesterone. The structures revealed a surprisingly small active site, with little conformational change associated with the binding of either compound. An unexpected peripheral binding site is identified, located above a phenylalanine cluster, which may be involved in the initial recognition of substrates or allosteric effectors.

CYP3A4 is a complex heme-containing enzyme that exhibits non–Michaelis-Menten kinetics and both homotrophic and heterotrophic cooperativity toward several substrates (*1*, *2*). For example, CYP3A4 exhibits atypical kinetic behavior in vitro with aflatoxin B1, amitriptyline, cabamazepine, proges-

terone, and diazepam (*3*, *4*). Studies have suggested that CYP3A4 has a noncatalytic effector site within the active-site cavity, capable of modulating function. In this model, the substrate and effector molecules occupy separate positions and only the substrate molecule has access to the reactive oxygen (*5*). These observations, combined with the vast chemical diversity of CYP3A4 substrates, have led to speculation that CYP3A4 has an adaptive active site capable of binding multiple compounds simultaneously (*6–10*). After examining the interaction of 10 probe substrates and finding that they could be divided into groups related to their size or chemical class, Kenworthy *et al.* (*11*) pro-

[1]Astex Technology, 436 Cambridge Science Park, Milton Road, Cambridge, CB4 0QA, UK. [2]Global Phasing Ltd., Sheraton House, Castle Park, Cambridge, CB3 0AX, UK.

*Present address: AstraZeneca, Research and Development, Charnwood, Bakewell Road, Loughborough, LE11 5RH, UK.
†To whom correspondence should be addressed. E-mail: h.jhoti@astex-technology.com

posed that three subpockets exist within the active site of CYP3A4. Additional studies have demonstrated the existence of multiple kinetically distinguishable conformations of CYP3A4, in the presence or absence of a substrate, effector, or inhibitor molecule(s) (12–14), which may contribute to allostery. External factors such as the binding of cytochrome b_5 and the presence of phospholipids (15) also affect the activity of CYP3A4 toward some compounds and may also stabilize particular conformations.

Understanding how this complex enzyme recognizes multiple, diverse chemical structures has been hampered by the lack of knowledge about the three-dimensional structure of CYP3A4. Homology models (16, 17) have limited use because of the substantial sequence diversity of P450s. The most widely used templates for modeling CYP3A4 are the P450 structures from *Bacillus megaterium* (P450 BM3) and *Saccharopolyspora erythraea* (P450 EryF), both of which share less than 25% sequence identity with CYP3A4. Furthermore, the sequence identity with human CYP2C9, for which the crystal structure was recently determined (18), is only 24%. We report crystal structures of CYP3A4, both unliganded and bound to either the inhibitor metyrapone or the substrate progesterone. The protein used in these studies was N-terminally truncated to fa-

cilitate crystallization and shown to be functionally active (19).

The ligand-free structure of CYP3A4 was solved to a resolution of 2.8 Å by a combination of multiple-wavelength anomalous dispersion and molecular replacement techniques (19) (tables S2 and S3). The overall structure of CYP3A4 conforms to the fold that is characteristic of the P450 superfamily, with a small, predominantly β-strand N-terminal domain and a larger, helical, C-terminal domain that contains the heme and the active site (Fig. 1A). The heme iron is ligated by a conserved cysteine (Cys[442]) and the propionates of the heme interact with the side chains of Arg[105], Trp[126], Arg[130], Arg[375], and Arg[440]. The heme is accessible to solvent by means of channels formed by β sheet 1, the B-C loop and the F-G region (Fig. 1B and fig. S1). There is a channel running from the active site to the outside of the protein formed by the B-C loop and the G′ helix, whereas another channel is formed by the F′ helix and the β sheet 1. The two channels are separated by the loop between the B′ and the C helices, a region well recognized to adopt different conformations dependent on the presence of a ligand (20). Although the heme of the protein sample used for crystallization was ferric and low spin,

x-rays are known to be capable of reducing heme iron, and thus its oxidation state in the structure is unknown (21). A low-spin ferric heme iron would be expected to have a ligand or water molecule in the sixth coordination position, but the structure reveals no ordered water bound in this position. A five-coordinate heme iron would require that the iron be displaced out of the plane of the coordinating porphyrin nitrogen atoms and toward the cysteine sulfur atom. However, given the resolution of the data, no conclusions can be reliably drawn regarding coplanarity of the heme iron. Another functionally important water molecule in other P450s is located between the highly conserved residues Ala[305] and Thr[309], which lie on the I helix and are close to the heme. In some P450 structures, the threonine residue interacts with this water molecule, which disrupts the hydrogen-bonding pattern of the secondary structure, resulting in a kink in helix I (22). Although a similar but less pronounced kink is observed in CYP3A4, there is no discrete density visible for this water molecule.

CYP3A4 has a number of features that differ from other P450 structures. A notably hydrophobic region is located around the loop following helix A″ (top left, Fig. 1A), a helix not observed in other P450 crystal structures. This region, along with the G′ helix and the loop between the G′ and G helices, which are also hydrophobic,

Fig. 1. (A) Overall fold of CYP3A4, colored from blue at the N terminus to green, to yellow, to red at the C terminus. The same coloring scheme is used for CYP3A4 in (B) and in Figs. 3 and 4 and for CYP2C9 in (C). The heme is depicted as a ball-and-stick model in the center of the molecule, flanked by helix I. Helix A″, not present in other P450 structures, is shown at the top left of the structure, and the shortened F helix is to the right of the heme. All figures were produced with AstexViewer version 2 (36). **(B)** View of the Phe-cluster of CYP3A4. The solvent-accessible

molecular surface of the active-surface cavity of CYP3A4 is depicted as a semitransparent green surface. The seven phenylalanine residues that make up the Phe-cluster are depicted as ball-and-stick models. The available volume of the active site is about 520 Å³ [calculated with the LIGSITE algorithm (28)]. The short F helix is shown at the right of the structure, and leads into a loop region that contains Phe[213], Phe[215], Phe[219], and Phe[220] before the F′ helix. The Phe-cluster is about 10 Å above the heme, which is depicted as a ball-and-stick model at the bottom of the structure. **(C)** View of the active-site cavity of CYP2C9. The solvent-accessible molecular surface of CYP2C9 (PDB entry 1OG2) is depicted as a semitransparent surface in purple. The available volume of the active site is about 600 Å³ [calculated with the LIGSITE algorithm (28)]. The F helix of CYP2C9 is longer than that of CYP3A4, and the relative position of secondary structural elements results in a narrowing of the active site toward the heme and a substantially different shape when compared with the active site of CYP3A4.
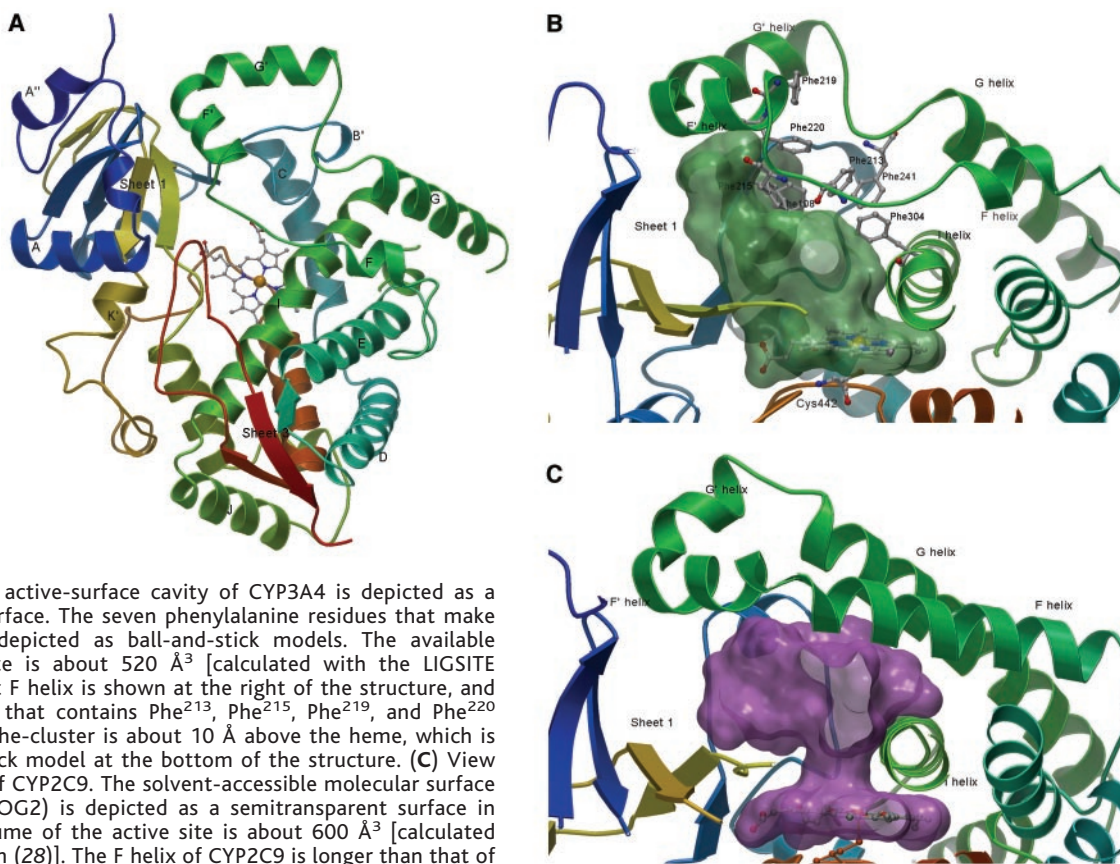
may mediate interaction with the microsomal membrane. Another unexpected feature of the CYP3A4 structure is the region following helix F (Fig. 1, A and B). This helix is markedly short compared with other P450 structures, and it leads into an ordered stretch of polypeptide chain that does not conform to any secondary structure motif. This region is located above and perpendicular to helix I and contains a number of residues that have been shown by site-directed mutagenesis to have a direct or indirect role in CYP3A4 function. For example, Leu[210], implicated in effector binding as well as the stereo- and regioselectivity profile of CYP3A4 (7, 23, 24), lies in this region with its side chain pointing away from the active site. Leu[211] and Asp[214], implicated in cooperativity (23, 25), also lie in this extended loop region and point away from the active site. In contrast, the side chains of two other residues shown by mutagenesis to have no role in cooperativity (23), Phe[213] and Phe[215], point toward the active site and, together with another five phenylalanine residues (Phe[108], Phe[219], Phe[220], Phe[241], and Phe[304]), form a "Phe-cluster" (Fig. 1B). No other structurally characterized P450 has such a cluster of phenylalanine residues in this region. Some of these other Phe-cluster residues have also been shown by mutagenesis studies to be involved in CYP3A4 activity. For example, Phe[304], which lies on the I helix, has a dual role in cooperativity, regioselectivity, and stereoselectivity (7, 26, 27), whereas the substitution of Phe[108] with a smaller or larger amino acid affected the metabolism of some substrates (7).

The Phe-cluster lies above the active site, with the aromatic side chains stacking against each other to form a prominent hydrophobic core. This region of the structure is highly ordered with an average B factor of 41 Å^2 for the residues in the cluster, compared with the average of 66 Å^2 over the entire structure. Furthermore, as a result of this aromatic clustering, the active site of CYP3A4 has a probe-accessible volume (28) of about 520 Å^3, substantially smaller than expected on the basis of its reported metabolism of large substrates. The overall volume of the CYP3A4 active site is also smaller than that of CYP2C9 (18), which is 670 Å^3, and

different in shape (Fig. 1, B and C). The variation in the active-site topologies is a consequence of the Phe-cluster, which is positioned on the top of the active site closer to the heme, with β sheet 1 farther from the heme compared with CYP2C9 (Fig. 1C). An overlay of CYP3A4 with the structures of P450 BM3 (29) and P450 EryF (30) (Fig. 2) shows the limitations of homology models based on these two bacterial structures. The position of the Phe-cluster results in a closed conformation of the active site of CYP3A4 that is similar to the closed conformation adopted by P450 EryF (30), but the spatial arrangement of the F-G loop and F helices is not conserved (Fig. 2). The heme of CYP3A4 has greater accessibility to the active site than does that of CYP2C9 (Fig. 1C), which could allow two substrate molecules to have access to the reactive oxygen, consistent with data indicating that CYP3A4 is able to bind and metabolize multiple substrate molecules simultaneously (6, 8). Conformational movement involving the Phe-cluster could also reposition phenylalanine residues, perhaps resulting in an extension of helix F and a larger active site. There is a precedent for such a mechanism from the CYP119 structure in which, although there is no analogous Phe-cluster, ligand binding resulted in an extension of helix F and movement of the G helix and the intervening F-G loop (31). To investigate whether conformational movement is a necessary prerequisite for ligand binding by CYP3A4, we determined the crystal structure in complex with the known inhibitor metyrapone. The metyrapone cocomplex structure was determined with the use of both soaking and, more importantly, cocrystallization techniques, which allowed conformational movement by the protein if required.

Contrary to expectations, the binding of metyrapone to CYP3A4 reveals essentially no conformational changes in the protein. Ultraviolet and visible spectroscopic data (32) indicate that in solution the inhibitor ligates the heme. This is consistent with the binding mode observed in the crystal structure, in which metyrapone is bound directly to the heme iron by means of the alkyl-pyridine nitrogen (Fig. 3), and exhibits good shape complementarity with the CYP3A4 active site (fig. S2). An alternative binding mode for metyrapone may involve the nitrogen atom of the keto-pyridine group, as observed previously in the cocomplex with bacterial P450 cam (33). However, the electron density for this CYP3A4 cocomplex indicates that the illustrated binding mode is predominant. The volume occupied by the metyrapone molecule is 230 Å^3, leaving sufficient space for additional molecules to bind within the active site.
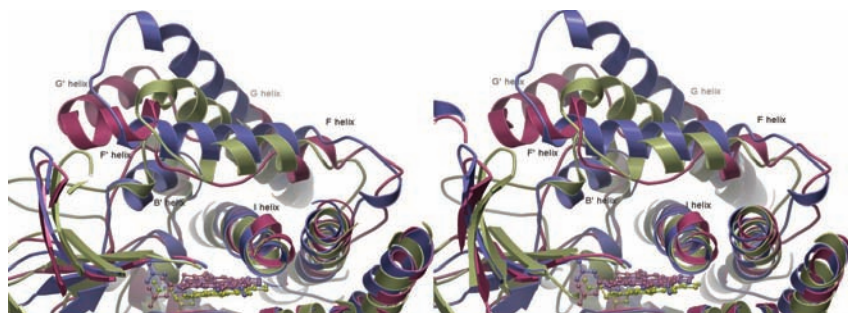

Fig. 2. Stereoimage showing the structural overlay of CYP3A4 in red, with P450 BM3 (molecule A of PDB entry 1BU7) in blue, and P450 EryF (PDB entry 1OXA) in yellow. The heme is depicted as a ball-and-stick model toward the bottom of the figure. The F helix of CYP3A4 (red) is notably shorter than the F helices of P450 BM3 (blue) or P450 EryF (yellow). CYP3A4 has two helices F′ and G′ that are absent in P450 BM3, P450 EryF, and other bacterial P450 structures.
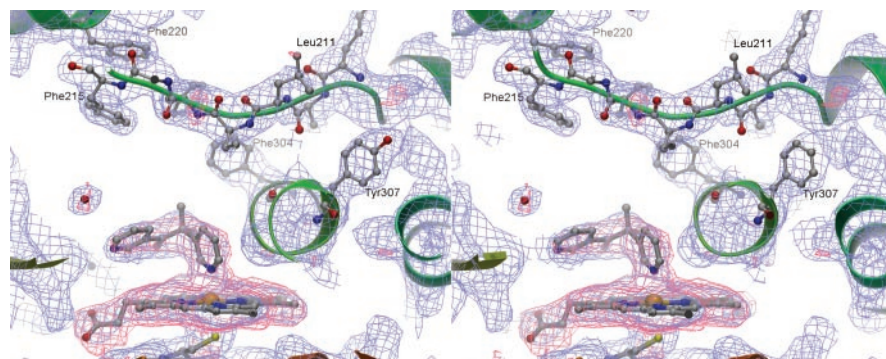

Fig. 3. Stereoview of metyrapone bound to the heme of CYP3A4. The metyrapone, heme, and surrounding residues are depicted as ball-and-stick models. The initial sigmaA-weighted Fo-Fc electron density map is shown in red contoured at the 2.0σ level, and the final sigma-weighted Fo-Fc electron density map, calculated omitting the heme, is shown in red contoured at the 2.0σ level, and then is shown in blue contoured at the 1.0σ level. The compound exhibits a single binding orientation and interacts with the protein by means of a pyridinyl group of metyrapone.

Fig. 4. Peripheral binding of progesterone to CYP3A4. The solvent-accessible surface of the progesterone molecule is shown in yellow at the top of the figure. A cross section of the solvent-accessible surface of CYP3A4 is shown in blue, revealing the active-site cavity of CYP3A4 in the center of the figure. Progesterone sits in a shallow pocket on top of the Phe-cluster of CYP3A4, separated from the heme by the F-G region of the molecule. Progesterone binds 17 Å above the heme iron and on the face of CYP3A4 most likely to interact with the membrane.
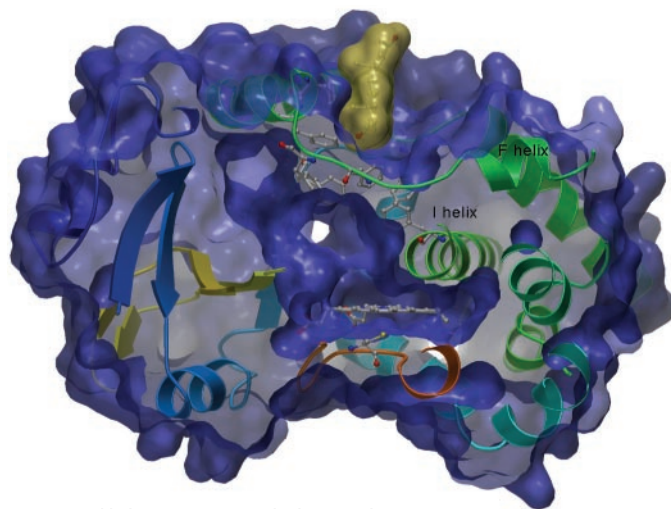
Although formation of the metyrapone complex is achieved with little conformational change, we anticipate that substantial protein movement, possibly involving the F and G helical regions and the Phe-cluster, would be required to permit the entry and accommodation of larger compounds. Movement of the B-C loop region could also open up the channel and increase the volume of the active site. To investigate this possibility and explore how CYP3A4 recognizes substrates, we determined the structure of the complex with progesterone by cocrystallization. Surprisingly, progesterone also induced very little conformational change in the protein and bound at a peripheral site very close to the Phe-cluster (Fig. 4) and some distance away (>17 Å) from the heme iron. The initial electron density maps indicated that one progesterone molecule bound in a single clearly defined orientation (fig. S3), as reflected by an average B factor for progesterone of 45 $Å^2$, compared with 61 $Å^2$ for the protein. The progesterone molecule interacts with the protein through a hydrogen bond between its acetyl oxygen and the amide nitrogen of Asp^{214}, packs against the side chains of Phe^{219} and Phe^{220} in the Phe-cluster (fig. S3), and appears fairly accessible to solvent. Although this binding site may be an artifact of crystallization, we argue that it is likely to have functional relevance. This is supported by the observation that the side chains of Leu^{211}, Phe^{213}, and Asp^{214} residues implicated in both progesterone homotrophic and heterotrophic cooperativity (5, 25) lie in the vicinity of this peripheral binding pocket. On the basis of these findings, we propose that the progesterone-binding site may be involved in the recognition of effector as well as substrate molecules and has a role in modulating cooperativity. Notably, although progesterone can be docked into the active site of CYP3A4 in an orientation consistent with its metabolism, there is no evidence of progesterone binding within the active-site cavity in this structure.

Several studies indicate that the F-G and B-C loops of P450, which constitute the putative substrate-access channel, could contribute to membrane binding and orient the substrate-access channel toward the membrane surface (34, 35). The location of the progesterone-binding site in the F-G region is consistent with a role in initial substrate recognition, because it lies along an appropriate route for compounds that move directly from the membrane into the active site of CYP3A4. Appropriate conformational movement of residues around the Phe-cluster could result in a substrate-access channel that stretched from this peripheral binding site to the heme group, providing a route for a compound to move from this initial recognition site to the active site. Such conformational movements in the Phe-cluster could be triggered by interaction with a physiological electron transfer partner such as cytochrome b_5 or cytochrome P450 reductase, or by a change in membrane properties to allow the substrate molecule to move from its surface position to a position in the active site that is suitable for metabolism. Future studies that investigate these potential mechanisms of molecular recognition and enzyme activity can now be guided by the crystal structure of CYP3A4.

References and Notes

1. F. P. Guengerich, *Annu. Rev. Pharmacol. Toxicol.* **39**, 1 (1999).
2. J. M. Hutzler, T. S. Tracy, *Drug Metab. Dispos.* **30**, 355 (2002).
3. Y. F. Ueng, T. Kuwabara, Y. J. Chun, F. P. Guengerich, *Biochemistry* **36**, 370 (1997).
4. G. E. Schwab, J. L. Raucy, E. F. Johnson, *Mol. Pharmacol.* **33**, 493 (1988).
5. T. L. Domanski *et al.*, *Biochemistry* **40**, 10150 (2001).
6. K. R. Korzekwa *et al.*, *Biochemistry* **37**, 4137 (1998).
7. K. K. Khan, Y. Q. He, T. L. Domanski, J. R. Halpert, *Mol. Pharmacol.* **61**, 495 (2002).
8. M. Shou *et al.*, *Biochemistry* **33**, 6450 (1994).
9. K. E. Kenworthy, S. E. Clarke, J. Andrews, J. B. Houston, *Drug Metab. Dispos.* **29**, 1644 (2001).
10. P. Lu *et al.*, *Drug Metab. Dispos.* **29**, 1473 (2001).
11. K. E. Kenworthy, J. C. Bloomer, S. E. Clarke, J. B. Houston, *Br. J. Clin. Pharmacol.* **48**, 716 (1999).
12. A. P. Koley, J. T. Buters, R. C. Robinson, A. Markowitz, F. K. Friedman, *J. Biol. Chem.* **270**, 5014 (1995).
13. A. P. Koley, J. T. Buters, R. C. Robinson, A. Markowitz, F. K. Friedman, *J. Biol. Chem.* **272**, 3149 (1997).
14. D. R. Davydov, J. R. Halpert, J. P. Renaud, H. G. Hui Bon, *Biochem. Biophys. Res. Commun.* **312**, 121 (2003).
15. K. H. Kim, T. Ahn, C. H. Yun, *Biochemistry* **42**, 15377 (2003).
16. D. F. Lewis, P. J. Eddershaw, P. S. Goldfarb, M. H. Tarbit, *Xenobiotica* **26**, 1067 (1996).
17. G. D. Szklarz, J. R. Halpert, *J. Comput. Aided Mol. Des.* **11**, 265 (1997).
18. P. A. Williams *et al.*, *Nature* **424**, 464 (2003).
19. Materials and methods are available as supporting material on *Science* Online.
20. T. L. Poulos, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 13121 (2003).
21. C. Nave, *Radiat. Phys. Chem.* **45**, 483 (1995).
22. D. C. Haines, D. R. Tomchick, M. Machius, J. A. Peterson, *Biochemistry* **40**, 13456 (2001).
23. G. R. Harlow, J. R. Halpert, *J. Biol. Chem.* **272**, 5396 (1997).
24. H. Wang *et al.*, *Biochemistry* **37**, 12536 (1998).
25. G. R. Harlow, J. R. Halpert, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 6636 (1998).
26. T. L. Domanski, J. Liu, G. R. Harlow, J. R. Halpert, *Arch. Biochem. Biophys.* **350**, 223 (1998).
27. J. C. Stevens *et al.*, *J. Pharmacol. Exp. Ther.* **290**, 594 (1999).
28. M. Hendlich, F. Rippmann, G. Barnickel, *J. Mol. Graph. Model.* **15**, 355 (1997).
29. I. F. Sevrioukova, H. Li, H. Zhang, J. A. Peterson, T. L. Poulos, *Proc. Natl. Acad. Sci. U.S.A* **96**, 1863 (1999).
30. J. R. Cupp-Vickery, T. L. Poulos, *Nature Struct. Biol.* **2**, 144 (1995).
31. S. Y. Park *et al.*, *J. Inorg. Biochem.* **91**, 491 (2002).
32. P. A. Williams *et al.*, data not shown.
33. T. L. Poulos, A. J. Howard, *Biochemistry* **26**, 8165 (1987).
34. J. Cosme, E. F. Johnson, *J. Biol. Chem.* **275**, 2545 (2000).
35. S. E. Graham-Lorence, J. A. Peterson, in *Physiological Functions of Cytochrome P450 in Relation to Structure and Regulation*, C. Jefcoate, Ed. (Jai Press, Greenwich, CT, 1995), pp. 57–79.
36. M. J. Hartshorn, *J. Comput. Aided Mol. Des.* **16**, 871 (2002).
37. We thank S. Rich, S. Maman, D. Cross, S. Kirton, and C. Murray for assistance with crystallization, characterization, and analysis; G. Williams for discussions; and F. W. Dahlquist for the provision of the pCWori+ expression vector. We acknowledge the European Synchrotron Radiation Facility for the provision of synchrotron radiation facilities and thank the industrial user group for assistance with data collection. Coordinates and structure factors of the ligand-free, metyrapone-, and progesterone-bound structures of CYP3A4 have been deposited in the Protein Data Bank (PDB) (entries 1W0E, 1W0F, and 1W0G).

NEW PRODUCTS

ESA

For more information
+44 1844-239381
www.esainc.com
www.scienceproductlink.org

ELSD HPLC IMPROVEMENT

New remote control software for the Chromachem Evaporative Light Scattering Detector (ELSD) simplifies detector operation and high-performance liquid chromatography (HPLC) system integration. The Chromachem ELSD provides universal detection of non-volatile compounds. Unlike HPLC ultraviolet or fluorescence detectors, the Chromachem ELSD does not require the presence or absence of chromophores, fluorophores, or electroactive groups for detection. The technique is fully compatible with both organic solvent–based and aqueous mobile phases, even at high or microbore flow rates (20 μl/min to 3 ml/min). The new software features single-point control from an external computer of all detection parameters and programmed methods, as well as manual and automated control of external events and real-time data monitoring. The software allows users to easily create, store, and recall operation methods and collected data files directly on a Windows-based personal computer. The Chromachem ELSD is suited for the detection of a wide range of weakly chromophoric compounds, including lipids, phospholipids, triglycerides, carbohydrates, polymers, pharmaceuticals, and combinatorial library compounds.

Eppendorf

For more information
+49 40-5 38 01-0
www.eppendorf.com
www.scienceproductlink.org

PEPTIDE TIPS

The PerfectPure C-18 Tip is a tool for desalting, purifying, and concentrating peptides for mass spectrometry. The tips' most significant advantage is the high degree of reproducibility of the flow-through characteristics, which is achieved with innovative manufacturing technology and strict quality control. The product specifications are tailored to meet the requirements of peptide mass fingerprinting, the identification of protein samples using mass spectrometry analysis. Alternative elution protocols, which may increase peptide recovery, are available for more complex matrices. The tip fits on most common pipet brands.

Guava Technologies

For more information
866-448-2827
www.guavatechnologies.com
www.scienceproductlink.org

CELL MIGRATION RESPONSE

Using Guava Express and the Guava PCA cell analysis system, researchers can quickly and easily study cell migration and screen for drug compounds that affect that process. The use of Guava Express in an actin polymerization assay can provide an early measure of chemotaxis, within seconds of exposure to the stimuli. The actin assay can be readily adopted to identify drug compounds that enhance or inhibit chemotaxis or to study pathways involved in chemotaxis.

C.B.S. Scientific

For more information
800-243-4959
www.cbsscientific.com
www.scienceproductlink.org

PCR WORKSTATIONS

The PCR Workstation offers microscope compatibility, accommodating both single and dual ocular microscopes. The workstation is designed to protect against polymerase chain reaction (PCR) contamination in two ways. Ultraviolet (UV) irradiation of the working area prior to use blocks replication of contaminating DNA sequences by causing adjacent pyrimidines to undergo dimeriz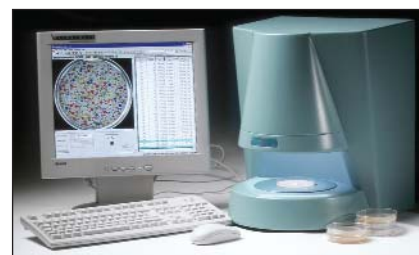ation. The workstation protects against cross contamination or airborne contamination by limiting exposure of the experimental set-up to the open lab environment. A small oval cut-out is removed from the tempered glass facia to accommodate the oculars. The opening is then covered with silicone, which permits transmittance of only negligible amounts of both UVA and UVB, and is cut to securely wrap the oculars. The silicone also protects the sterility of the microscope stage area by preventing contaminating DNA sequences from falling through the ocular cut-out.

Synbiosis

For more information
+44 (0) 1223 727125
www.synbiosis.com
www.scienceproductlink.org

DIGITAL CAMERAS FOR MICROBIOLOGY

A range of high-resolution, color digital cameras can be connected to a ProtoCOL automated colony counting and zone sizing system as a cost-effective method to automatically analyze samples via a microscope or on bioassay plates. The cameras' color reproduction and high resolving power make them suitable for accurately imaging colonies, zones, or plaques. The cameras come equipped with the latest Fire-Wire technology to allow microbiologists to see real-time images and save time when analyzing samples. The cameras enable users to read a wide range of media and plates types.

Ambion

For more information
800-888-8804
www.ambion.com
www.scienceproductlink.org

RNA ISOLATION

The RNAqueous-4PCR Kit isolates RNA without genomic DNA contamination from samples as small as 1 mg or 100 cells. The rapid, phenol-free isolation procedure yields total RNA that is treated with deoxyribonuclease (DNase) in an optimized reaction buffer. The DNase and divalent cations are removed with a novel, nontoxic DNase Removal Reagent providing total RNA ready for reverse transcription-polymerase chain reaction analysis (RT-PRC). It is designed for both traditional end-point RT-PCR analysis and real-time PCR applications.

Stoelting

For more information
630-860-9700
www.stoeltingco.com
www.scienceproductlink.org

LITERATURE

Stoelting 2004–2005 is a physiology research instruments catalog for neuroscience, physiology, and general laboratory equipment. This 112-page, illustrated catalog lists hundreds of products from surgical needs to stereotaxic instruments, including more than 50 new products.

Newly offered instrumentation, apparatus, and laboratory materials of interest to researchers in all disciplines in academic, industrial, and government organizations are featured in this space. Emphasis is given to purpose, chief characteristics, and availability of products and materials. Endorsement by *Science* or AAAS of any products or materials mentioned is not implied. Additional information may be obtained from the manufacturer or supplier by visiting www.scienceproductlink.org on the Web, where you can request that the information be sent to you by e-mail, fax, mail, or telephone.